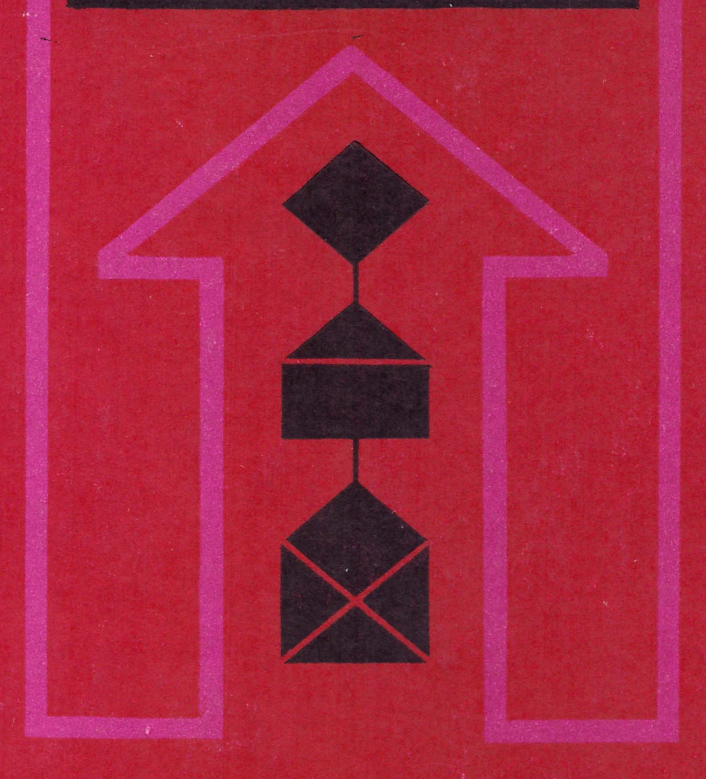


В. А. ИГНАТОВ

ТЕОРИЯ ИНФОРМАЦИИ И ПЕРЕДАЧИ СИГНАЛОВ



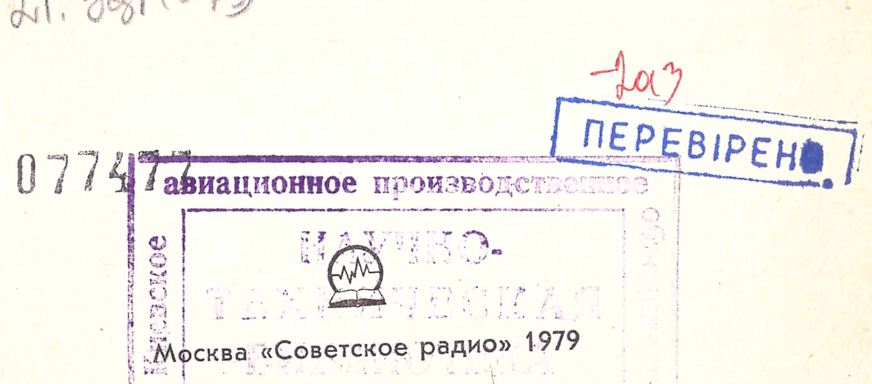
В. А. ИГНАТОВ • ТЕОРИЯ ИНФОРМАЦИИ И ПЕРЕДАЧИ СИГНАЛОВ



В. А. ИГНАТОВ

ТЕОРИЯ ИНФОРМАЦИИ И ПЕРЕДАЧИ СИГНАЛОВ

Утверждено УУЗ МГА СССР в качестве учебника для студентов вузов гражданской авиации



Москва «Советское радио» 1979

ББК 32.81

И 26

УДК 621.391

Игнатов В. А. Теория информации и передачи сигналов: Учебник для вузов. — М.: Сов. радио, 1979. — 280 с. с ил.

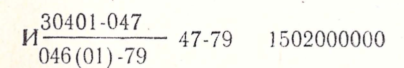
Систематизированно излагаются основные положения теории информации и передачи сигналов, дана общая характеристика задач этой теории. Приводятся методы математического описания сообщений, сигналов, помех и каналов связи, методы управления информационными параметрами сигналов, информационные характеристики источников сообщений, сигналов, помех и каналов, помехоустойчивость передачи дискретных и непрерывных сообщений, корректирующее кодирование. Рассмотрены принципы уплотнения линий связи, оценки и повышения эффективности передачи информации, особенности управления информационными потоками в сетях. В каждом разделе формулируются контрольные вопросы и выводы, приводятся численные примеры, способствующие лучшему усвоению материала.

Книга является учебником для студентов радиотехнических факультетов вузов гражданской авиации. Она может быть использована студентами других вузов и специалистами в области электросвязи, технической кибернетики и вычислительной техники.

Табл. 10, рис. 66, библ. 18 назв.

Рецензенты: кафедра авиационной связи Рижского института инженеров гражданской авиации (заведующий кафедрой канд. техн. наук В. А. Анисимов); докт. техн. наук проф. В. И. Васильев (Московский институт инженеров гражданской авиации).

Редакция литературы по вопросам космической радиоэлектроники



И $\frac{30401\text{-}047}{046(01)\text{-}79}$ 47-79 1502000000

ПРЕДИСЛОВИЕ

Учебник по курсу «Теория информации и передачи сигналов» предназначен для студентов, которые специализируются в области технической эксплуатации наземного и бортового авиационного радиооборудования. В основу учебника положены лекции, читаемые автором студентам радиотехнического факультета Киевского института инженеров гражданской авиации.

В учебнике систематизированно излагаются основы теории анализа, синтеза и оптимизации таких объектов информационной техники, как системы и сети электросвязи, информационно-измерительные системы и комплексы, вычислительные системы и комплексы, информационно-логические системы, подсистемы информационного обеспечения автоматизированных систем управления и т. п. Информационные системы различного назначения широко применяют в гражданской авиации для выполнения важных и ответственных функций по обеспечению безопасности и регулярности полетов. Поэтому курс «Теория информации и передачи сигналов» является одним из основных при подготовке радиоинженеров гражданской авиации.

При написании учебника учтены опыт и методика изложения материалов всех учебников этого направления. Автор признателен |Н. Д. Босому|, С. М. Пауку, В. С. Козлову, Э. А. Корнильеву, советы которых способствовали улучшению книги. Большую помощь автору в работе над учебником оказали В. М. Чуприн, М. И. Лукина, В. Н. Юдин, И. Ф. Шевченко, А. М. Николенко. Автор выражает благодарность рецензентам В. И. Васильеву, В. А. Анисимову, А. Н. Левину за полезные замечания.

ВВЕДЕНИЕ

В.1. Предмет, место и роль теории информации и передачи сигналов. Научно-технический прогресс сопровождается интенсивным ростом объемов информации, необходимой для управления промышленностью, сельским хозяйством, транспортом и другими отраслями народного хозяйства. Об этом можно судить, к примеру, по увеличению объема информации, связанной с работой современного аэропорта гражданской авиации, которое пропорционально квадрату увеличения объема перевозок [8].

Важную роль в решении задач, поставленных XXV съездом КПСС, намеченных в Программе КПСС и сформулированных в Постановлениях партии и правительства, играют такие виды передачи информации, как радиовещание, телевидение, телефония, факсимильная связь, передача данных, в частности, в распространении научных, технических, экономических и культурных знаний; в удовлетворении потребности людей в общении.

В теории информации и передачи сигналов *под информацией* понимают совокупность сведений о каких-либо событиях, процессах, объектах, явлениях и т. п., рассматриваемых в аспекте их передачи в пространстве и во времени [17]. В более общем смысле «информация — это содержание связи между материальными объектами, проявляющееся в изменении состояний этих объектов» [17]. «*Информация* выступает как свойство объектов порождать многообразие состояний, которые посредством отражения передаются от одного объекта к другому...» [18].

Информацию передают в виде сообщений. Сообщением называют информацию, выраженную в определенной форме и предназначенную для передачи от источника информации к адресату. Примерами сообщений служат тексты телеграмм и фототелеграмм, речь, музыка, телевизионное изображение, данные на выходе ЦВМ, команды в системе управления воздушным движением и т. д. *Отправителями и получателями сообщений* могут быть как люди, так и технические устройства, которые накапливают, хранят, регистрируют, преобразуют, передают и принимают информацию.

Сообщения передают с помощью сигналов — материальных носителей информации. Сигналом может быть любой физический процесс, параметры которого могут отображать сообщение. Основными сигналами, которые рассматривают в теории информации и передачи сигналов, являются электрические сигналы. Однако методы этой теории применимы и к другим сигналам, например гидроакустическим и оптическим.

Предметом теории информации и передачи сигналов является изучение процессов накопления, измерения, переработки, хранения,

преобразования, передачи и приема информации. В узком смысле предметом теории информации и передачи сигналов считают изучение тех процессов, которые имеют место при передаче информации на расстояние посредством электрических сигналов.

Теория информации и передачи сигналов тесно связана с технической кибернетикой. Важными идеями кибернетики, как известно, являются идея о наличии обратной связи в живых и неживых системах, идея о передаче информации и идея об автоматических вычислительных машинах, способных выполнять функции, аналогичные мышлению человека. Идея о передаче информации является одной из центральных идей кибернетики. Управление процессами без получения и обработки информации невозможно, а получение информации без цели управления в большинстве случаев бессмысленно. Следовательно, понятия управление и информация неразрывно связаны. В кибернетике большее внимание уделяется проблемам управления, а в теории информации и передачи сигналов — проблемам передачи информации.

Высокие требования к качеству передачи информации особенно характерны для систем гражданской авиации, которая является одной из быстро развивающихся отраслей народного хозяйства. В настоящее время все большие средства вкладываются в развитие производства и интенсивной эксплуатации авиационной техники, строительство и реконструкцию крупных международных и магистральных аэропортов, систем межконтинентальной и ближней навигации, аэровокзальных комплексов, высокопроизводительных погрузочно-разгрузочных систем, автоматизированных систем управления с использованием ЦВМ и т. д. Непрерывно растут скорости, интенсивности, высоты, дальности полетов воздушных судов гражданской авиации.

За последние тридцать лет на порядок увеличилось количество приборов для контроля и управления полетами самолетов, в 2—3 раза сократилось время для принятия и выполнения решения на борту самолета. Усложнился и возрос темп радиообмена с землей.

Наблюдаются следующие основные тенденции развития информационной техники гражданской авиации: расширение областей применения и выполняемых функций, освоение новых диапазонов, усиление взаимосвязей между отдельными видами информационных систем; увеличение удельного веса устройств обработки информации на борту; повышение сложности систем наряду с ростом требований к качеству их функционирования; повышение требований к квалификации радиоинженеров; системный подход к анализу, синтезу и оптимизации объектов информационной техники на основе вероятностных и статистических методов.

Курс «Теория информации и передачи сигналов» является основой для развития инженерных методов расчета и проектирования конкретных объектов информационной техники в специальных курсах, поэтому значение и роль этой дисциплины в системе подготовки радиоинженеров трудно переоценить.

В.2. Классификация объектов информационной техники. По функциональному признаку можно выделить следующие основные классы объектов информационной техники:

— сети и системы электросвязи (телеграфные, телефонные, телевизионные сети и системы, системы радиовещания, системы передачи данных и т. п.);

— информационно-измерительные системы и комплексы (радиолокационные, радионавигационные, радиотехнические системы автоматического самолетовождения, радиотелеметрические и т. п.);

— системы преобразования информации (цифровые и аналоговые вычислительные машины, аналого-цифровые и цифроаналоговые преобразователи и т. п.);

— подсистемы информационного обеспечения автоматизированных систем управления;

— информационно-поисковые системы;

— системы хранения информации;

— сети и системы экспериментального наблюдения или исследования (астрономические, физические, медицинские и т. п.).

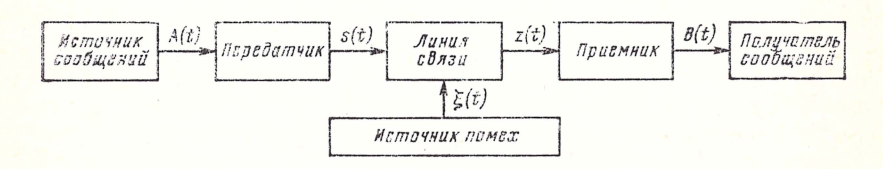


Рис. В.1. Структурная схема одноканальной системы передачи информации

Наибольшее развитие методы теории информации и передачи сигналов получили для анализа и синтеза систем электросвязи, так как именно с задач анализа простейших систем электросвязи началась сама теория информации. Это направление теории информации и передачи сигналов называют статистической теорией связи. В настоящее время методы теории информации и передачи сигналов все более успешно применяют для анализа и синтеза и других объектов информационной техники.

В.3. Цель, задачи и методы теории информации и передачи сигналов. Для рассмотрения цели, задач и методов теории информации и передачи сигналов используем обобщенные операторные уравнения передачи и обработки информации в одноканальных и многоканальных системах, информационных узлах, сетях распределения, накопления и передачи информации.

Простейшая структурная схема одноканальной системы передачи информации в одном направлении представлена на рис. В.1. Исходное сообщение $A(t)$ с выхода источника сообщений поступает на вход передатчика, назначение которого — преобразовывать сообщения в электрические сигналы $s(t)$, удобные для передачи по линии связи. Роль *линии связи* может играть любая физическая

среда (воздух, вода, провода и т. п.), способная пропускать сигналы. В приемнике принятые из линии связи электрические сигналы $z(t)$ преобразуются в копии сообщений $B(t)$, передаваемые получателю сообщений. Копия сообщения $B(t)$ в той или иной мере соответствует оригиналу $A(t)$.

Искажения сообщений в значительной степени обусловлены помехой $\xi(t)$. *Под помехой* понимают все те воздействия, которые искажают передаваемые сигналы и приводят к тому, что принятые сигналы $z(t)$ отличаются от переданных $s(t)$. Совокупность передатчик, приемник, линия связи с источником помех — называют каналом связи. Источник сообщений, канал связи и получатель сообщений образуют *одноканальную систему передачи информации* (СПИ).

Для вывода обобщенного операторного уравнения, описывающего функционирование одноканальной СПИ, последовательно рассмотрим конкретные операции преобразования сообщений и сигналов, которые имеют место, например, при передаче телеграфного текста по проводным каналам связи. Сообщение $A(t)$ в виде телеграммы поступает на вход передатчика, который выполняет операцию кодирования (см. § 1.6). При кодировании последовательность букв и цифр преобразуется в последовательность электрических импульсов $s(t)$. Характер $s(t)$ определяется видом кода. Следовательно, в результате кодирования получают электрический сигнал $s(t)=K[A(t), s_к]$, где $s_к$ — вектор, компонентами которого являются элементарные кодовые сигналы.

При передаче последовательности импульсов $s(t)$ по линии связи они искажаются из-за воздействия помехи $\xi(t)$. Эту операцию воздействия помех на полезный сигнал обозначим через L, тогда сигнал на входе приемника можно представить так: $z(t)=L\{K[A(t), s_к], \xi(t)\}$.

После декодирования D приемником принятых сигналов получают копию сообщения

$$B(t)=D[z(t)]=D\{L[K[A(t), s_к], \xi(t)]\}. \qquad (В.1)$$

Уравнение (В.1) описывает функционирование одноканальной СПИ в операторной форме. Как видим, для анализа и синтеза даже простейших одноканальных СПИ необходимо уметь математически описывать сообщения, сигналы, помехи, а также и различные операции преобразования сообщений и сигналов (см. Приложение).

Перейдем к рассмотрению операторного уравнения для многоканальной СПИ. *Многоканальная СПИ* — это совокупность технических устройств, обеспечивающих одновременную и взаимно независимую передачу по одной общей линии связи сообщений от многих отправителей [8]. Структурная схема многоканальной СПИ с частотным разделением сигналов представлена на рис. В.2. В результате кодирования сообщения в i-м канале и модуляции на выходе i-го передатчика образуется сигнал

$$s_i(t)=M_i\{K_i[A_i(t), s_{кi}], x_i\},$$

где x_i — поднесущая, используемая при модуляции. Канальные сигналы $s_i(t)$ поступают в устройство уплотнения линии связи, где в соответствии с оператором уплотнения U из них и несущей x образуется многоканальный линейный сигнал $s(t)=U[s_i(t), x]$.

В линии сигнал $s(t)$ взаимодействует с помехой $\xi(t)$. В результате сигнал на выходе линии связи $z(t)=L[s(t), \xi(t)]$. Из сигнала $z(t)$ устройство разделения канальных сигналов в соответствии с оператором разделения U^{-1} выделяет принятые канальные сигналы $z_i(t)$. Демодуляция $D_{\text{м}i}$ и декодирование $D_{\text{к}i}$ сигнала $z_i(t)$ приемником позволяют получить копию $B_i(t)$.

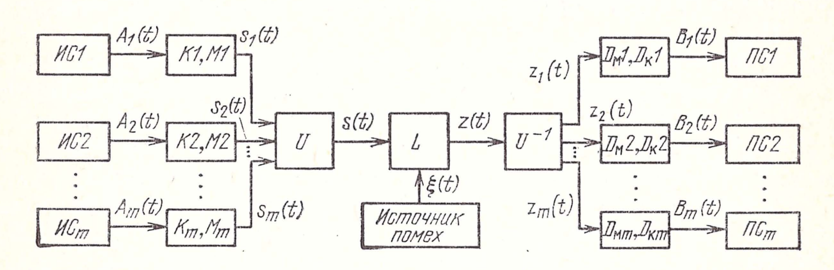


Рис. В.2. Структурная схема многоканальной системы передачи информации

Операторное уравнение, описывающее в целом работу многоканальной СПИ, имеет вид

$$B_i(t)=\{D_{\text{к}i}[D_{\text{м}i}[U^{-1}[L[U[M_i[K_i[A_i(t), s_{\text{к}i}], x_i], x], \xi(t)]]]]\}. \quad \text{(В.2)}$$

Следовательно, при анализе и синтезе многоканальной СПИ необходимо уметь математически описывать сообщения, сигналы, помехи и не менее семи операций преобразования сообщений, сигналов и помех.

Узел связи (информационный узел) предназначен для приема, хранения, преобразования и передачи информации, является еще более сложной системой. Помимо приемных и передающих устройств многоканальных систем, он содержит специализированные ЦВМ, выполняющие операции выбора кратчайшего пути ($V_{\text{п}}$), соблюдения системы приоритетов ($P_{\text{пр}}$) и накопления информации ($N_{\text{ин}}$) при отсутствии свободных каналов. На рис. В.3 показана упрощенная структурная схема информационного узла в режиме приема линейного сигнала по j-й линии связи и передачи линейного сигнала по k-й линии связи информационной сети. Индексом j помечены входные характеристики информационного узла, а индексом k — выходные, m — число каналов в j-й линии связи, q — число каналов в k-й линии связи.

Принципиальной особенностью информационного узла является то, что он выполняет сложные логические операции типа $V_{\text{п}}$, $P_{\text{пр}}$ и $N_{\text{ин}}$, реализовать которые можно на элементах цифровой электронной вычислительной техники. Эти операции формально описываются на языке теории алгоритмов, поэтому для анализа

и синтеза информационных узлов необходимо применять методы этой теории. Здесь, как видим, методы теории информации и передачи сигналов тесно переплетаются с методами вычислительной техники.

Сеть электросвязи (сеть распределения, накопления и передачи информации) является еще более сложным объектом информационной техники. Она представляет конечную совокупность информационных узлов, соединенных линиями связи. Описывать такие

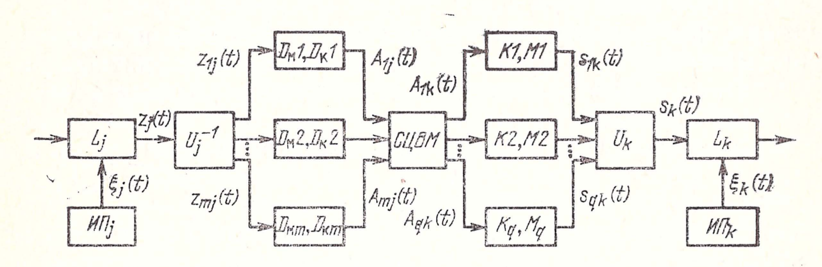


Рис. В.3. Упрощенная структурная схема информационного узла

сети удобно с помощью графов. На рис. В.4 показан пример графа информационные сети. Вершины графа u_i, $i=\overline{1,\ n}$, отражают информационные узлы, дуги — линии связи, через α_{ij}, β_{ji} обозначены вектор-параметры линий связи, координатами которых являются пропускная способность, интенсивность потока сообщений, стоимость и т. п. Стрелки указывают направления передачи информации.

Информационную сеть обычно рассматривают как сетевую систему массового обслуживания [18], поэтому основными методами анализа и синтеза сетей являются методы теории графов и теории массового обслуживания [18]. Организация оптимального управления потоками информации в сетях требует привлечения методов теории оптимального управления [18].

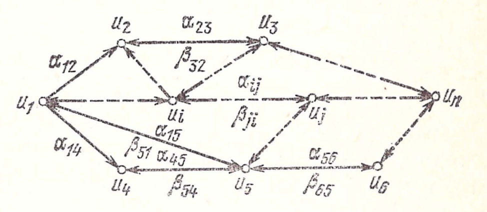


Рис. В.4. Граф информационной сети

Рассмотрение объектов информационной техники и операторных уравнений, описывающих в общем виде их работу, позволяет сформулировать *основную цель теории информации и передачи сигналов* — создание общих теоретических принципов и методов статистического анализа, синтеза и оптимизации объектов информационной техники.

Сущность задач анализа заключается в том, что необходимо определить, как влияют структура и параметры тех или иных известных операторов преобразований, сообщений, сигналов и помех

на характеристики качества передачи информации. *Сущность задач синтеза* заключается в том, что необходимо найти такие структуры и параметры операторов преобразований, сигналов, которые обеспечивают требуемое качество передачи информации. Если обеспечивается экстремум функционала качества, порождаемого этими операторами, то решается задача ***оптимального синтеза:*** выбора оптимальной структуры или оптимальных параметров операторов. Тогда говорят о структурной или параметрической оптимизации синтеза объекта.

Операторы реализуют с помощью аналоговых и дискретных функциональных преобразователей электрических сигналов. Поэтому задачи анализа, синтеза и оптимизации сводятся к анализу, синтезу и оптимизации структурных схем или параметров этих преобразователей. Определения преобразования, оператора, функционала и других, используемых в дальнейшем понятий функционального анализа даны в Приложении.

Теория информации и передачи сигналов достигла такого уровня развития, что уже позволяет решать задачи анализа, синтеза и оптимизации относительно простых информационных устройств и систем, таких как приемники, передатчики, модуляторы, демодуляторы, кодеры, декодеры и др. [6, 7, 12, 13, 18]. Методы оптимального синтеза объектов информационной техники интенсивно развиваются на основе системного подхода [18], позволяющего учитывать не только технические, но и экономические их характеристики.

Основными классами задач теории информации и передачи сигналов являются:

1. Анализ и синтез сигналов и помех.
2. Анализ и синтез каналов передачи информации.
3. Анализ информационных характеристик источников сообщений, сигналов, помех и каналов.
4. Анализ и синтез помехоустойчивых методов передачи информации.
5. Анализ и синтез корректирующих кодов (кодов, обнаруживающих и исправляющих ошибки).
6. Создание общих методов анализа и синтеза конкретных объектов информационной техники различных классов (см. п. В.2).

В теории информации и передачи сигналов используют методы функционального анализа; теории вероятностей и математической статистики; теории случайных функций и случайных процессов; статистической радиотехники; теории оптимальных статистических решений; математического программирования (методы отыскания экстремумов функций и функционалов при наличии ограничений) и др. [1—16].

Следовательно, теория информации и передачи сигналов носит комплексный характер, ее изучение требует хорошей математической подготовки.

В.4. Связь теории информации и передачи сигналов с другими дисциплинами. Теория информации и передачи сигналов излагает-

ся на основе знаний, полученных студентами при изучении высшей математики, физики, теории вероятностей и математической статистики, радиотехнических цепей и сигналов, статистической радиотехники, радиопередающих и радиоприемных устройств, вычислительной техники, импульсных устройств и других дисциплин.

В свою очередь, она служит теоретическим фундаментом для изучения таких дисциплин, как теория сетей электросвязи, системы электросвязи, теоретические основы радиолокации, теоретические основы радионавигации и управления воздушным движением, системы и устройства связи аэропортов, системы проводной связи, системы радиосвязи, системы радионавигации и радиолокации аэропортов гражданской авиации, связное оборудование летательных аппаратов; оптимальные сети связи и ряда других.

Поэтому курс «Теория информации и передачи сигналов» является одним из фундаментальных в подготовке радиоинженеров гражданской авиации.

В.5. Основные этапы развития теории информации и передачи сигналов. Первыми фундаментальными исследованиями, по существу породившими теорию информации и передачи сигналов, являются работы В. А. Котельникова по оптимальным методам приема сигналов на фоне помех (1946 г.) и работы К. Шеннона по обоснованию строгого измерения количества информации и оптимальным способам кодирования, обеспечивающим предельно достижимые скорость передачи информации по каналам и вероятность появления ошибок (1948 г.).

Однако элементы теории информации и передачи сигналов уже рассматривались в некоторых более ранних работах Р. Хартли по измерению количества информации (1928 г.), В. А. Котельникова о пропускной способности «эфира» и проволоки в электросвязи (1933 г.), Д. В. Агеева по основам теории линейной селекции сигналов (1935 г.) и других авторов.

Дальнейшему развитию теории передачи информации способствовало появление адекватного реальным процессам математического аппарата теории случайных функций и теории статистических решений. Этот аппарат позволил установить строгие количественные соотношения в теории передачи информации и сделать ее точной наукой. К концу 40—50-х гг. уже были хорошо известны ставшие классическими работы А. Я. Хинчина (1938 г.) по теории корреляции стационарных случайных функций, Н. Н. Колмогорова (1941 г.) и Н. Винера (1949 г.) по интерполированию и экстраполяции стационарных случайных последовательностей, А. Вальда (1950 г.) по статистическим решающим функциям.

По мере роста числа статей по теории информации и передачи сигналов и числа специалистов, применяющих вероятностные и статистические методы в таких задачах, как передача информации, автоматическое управление, радиотехнических и других, росло и число монографий, книг и учебников. Впервые монографии в этой области появились в 1957 г. (Левин Б. Р. Теория случайных процессов и ее применение в радиотехнике, Пугачев В. С. Теория случайных

функций и ее применение к задачам автоматического управления). В США в 1958 г. появилась монография В. Б. Давенпорта и В. Л. Рута «Введение в теорию случайных сигналов и шумов», в 1960 г.—двухтомная монография Д. Миддлтона «Введение в статистическую теорию связи». Позднее вышли в свет и учебники по теории передачи сигналов М. В. Назарова, Б. И. Кувшинова и О. В. Попова (1970 г.), А. Г Зюко и Ю. Ф. Коробова (1972 г.), Д. Д. Кловского (1973 г.), а также написанный с современных позиций учебник И. С. Гоноровского «Радиотехнические цепи и сигналы» (1976 г.).

Основными разделами теории информации и передачи сигналов вначале были теория структуры сигналов, теория информации и теория помехоустойчивости. По мере дальнейшего развития теории и получения новых результатов в таких направлениях, как математическое описание каналов передачи информации, создание методов помехоустойчивого и эффективного кодирования, уплотнения линий связи, распределения, накопления и передачи информации в сетях, а также внедрения методов теории информации в информационно-измерительную технику, все большее внимание уделяется вопросам системного оптимального синтеза объектов информационной техники. В настоящее время эти направления находятся на переднем рубеже научных исследований [5—16, 18].

Значительный вклад в развитие отдельных разделов теории информации и передачи сигналов внесли А. Н. Колмогоров, А. Я. Хинчин, В. А. Котельников, А. А. Харкевич, К. Шеннон, Н. Винер, Д. Миддлтон, Р. М. Фано, У. Питерсон, Р. Л. Добрушин, Л. Ф. Бородин, Д. Е. Вакман, Л. М. Финк, Л. С. Гуткин, Б. Р. Левин, С. Е. Фалькович, Р. Л. Стратонович, А. Г. Зюко, Н. Л. Теплов, Л. П. Пуртов и многие другие ученые.

В.6. Архитектоника учебника и методические указания по его использованию. Учебник включает предисловие, введение, десять глав основного текста, заключение, список литературы, Приложение и предметный указатель. Все разделы глав включают численные примеры и сопровождаются контрольными вопросами. Отличительной особенностью учебника является наличие гл. 1, 9, 10. Назначение гл. 1 — подготовить читателя к систематизированному восприятию курса «Теория информации и передачи сигналов», а 9 и 10 глав — для восприятия курсов по многоканальным системам и сетям связи. В этих главах дается общая характеристика и обсуждаются взаимосвязи основных задач теории информации и передачи сигналов, теории многоканальных систем и теории сетей электросвязи. Такая архитектоника обусловлена соображениями методического характера. Для первого чтения студентам, а также тем лицам, которые не специализируются в области передачи информации, а хотят иметь лишь общее представление о теории информации и передачи сигналов, достаточно ознакомиться с Приложением, предисловием, введением, гл. 1, 9, 10, выводами гл. 2—8 и заключением. Этот же материал рекомендуется для установочных лекций студентам заочного обучения.

Указанные материалы создают у читателей необходимые общие представления о задачах и методах теории информации и передачи сигналов, теории многоканальных систем, теории сетей электросвязи. Выводы к главам позволяют, во-первых, более четко систематизировать эти представления, во-вторых, дают возможность студентам приобщаться к творческой переработке информации и самостоятельно делать краткие выводы по изучаемым разделам курса, что полезно при использовании других учебников и при работе с научно-технической литературой, и, в-третьих, выводы вместе с примерами и контрольными вопросами помогают студентам проверить, насколько усвоен материал.

Выбранная методика изложения приводит к определенной избыточности. Однако опыт показывает, что на это идти целесообразно, так как существенно повышается эффективность обучения. Материалы гл. 2—9 рекомендуется использовать в той или иной мере в лекциях, на практических и лабораторных занятиях в соответствии с конкретными рабочими программами кафедр. Они также необходимы студентам при самостоятельном углубленном изучении курса.

Контрольные вопросы

1. Дайте определение понятия информации.
2. Что такое сообщение?
3. Какая связь существует между сигналом и сообщением?
4. Определите предмет теории информации и передачи сигналов.
5. Каковы основные тенденции развития информационной техники гражданской авиации?
6. Как классифицируют объекты информационной техники по функциональному признаку?
7. Как составляют операторные уравнения, описывающие работу систем передачи информации? Запишите и поясните операторное уравнение для одноканальной системы.
8. Что понимают под линией, каналом, системой передачи информации?
9. Укажите основную цель теории информации и передачи сигналов?
10. Назовите основные задачи теории информации и передачи сигналов.
11. В чем сущность задач анализа и синтеза систем передачи информации?
12. Какие методы использует теория информации и передачи сигналов?
13. Как связана теория информации и передачи сигналов с другими дисциплинами?
14. Укажите основные этапы развития теории информации и передачи сигналов.

Глава 1

ОБЩАЯ ХАРАКТЕРИСТИКА ЗАДАЧ ТЕОРИИ ИНФОРМАЦИИ И ПЕРЕДАЧИ СИГНАЛОВ

1.1. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ СООБЩЕНИЙ, СИГНАЛОВ И ПОМЕХ

Сообщения и соответствующие им сигналы могут быть дискретными или непрерывными. *Дискретное сообщение* представляет конечную последовательность отдельных символов (букв), длительность этой последовательности ограничена. Типичным примером дискретного сообщения является телеграмма. Дискретные сообщения характерны для телеграфии, передачи данных и телеметрии. Для преобразования дискретного сообщения в сигнал необходимо выполнить, как уже отмечалось во введении, операцию кодирования сообщения (см. § 1.6).

Непрерывное сообщение описывается непрерывной функцией времени. Примерами непрерывных сообщений служат речь, музыка, телевизионное изображение. С помощью специальных устройств непрерывные сообщения преобразуются в электрические непрерывные сигналы. Например, если сообщением служит речь, то микрофон преобразует звуковые колебания воздушной среды в электрические колебания. Электрические сигналы передаются в место приема или служат модулирующим сигналом для высокочастотного колебания.

Существуют системы электросвязи, предназначенные для передачи непрерывных сообщений дискретными методами. В этих системах непрерывные сигналы, соответствующие непрерывным сообщениям, с помощью операций дискретизации во времени и квантования по уровню (см. § 3.6) преобразуют в дискретные, которые затем передают дискретными методами. В месте приема из принятых дискретных сигналов восстанавливают переданные непрерывные сигналы.

При математическом описании сообщений *формирование дискретных сообщений* рассматривают как последовательный случайный выбор того или иного символа из алфавита источника сообщений, т. е. как формирование случайной дискретной последовательности. Формирование непрерывных сообщений рассматривают как выбор реализаций непрерывной случайной функции. Поэтому основными математическими моделями дискретных сообщений являются дискретные случайные последовательности, а непрерывных сообщений — непрерывные случайные процессы [15]. Построение

математических моделей сообщений осуществляется вероятностными методами, а оценка параметров моделей выполняется методами математической статистики.

Математической моделью называют систему математических соотношений, описывающих изучаемый процесс или явление. Построение математических моделей называют математическим моделированием [18]. Для моделирования выбирают подходящие математические средства — алгебраические, дифференциальные или интегральные уравнения, теорию множеств, функциональный анализ, математическую логику, теорию вероятностей, теорию случайных процессов и др. Математическими моделями электрических сигналов, отображающих сообщения, и помех также являются случайные последовательности и процессы, поэтому математическое описание как полезных сигналов, так и вредных (помех) осуществляется одними и теми же методами.

В основе математического описания сообщений, сигналов и помех лежат методы теории вероятностей, теории случайных функций и математической статистики. *Целью математического описания* является разработка математических моделей сообщений, сигналов и помех, необходимых для анализа, синтеза и оптимизации объектов информационной техники. Математические модели позволяют анализировать свойства сообщений, сигналов и помех, а также синтезировать сигналы с требуемыми свойствами.

Все реальные сигналы и помехи являются случайными. Несмотря на это, в теории информации и передачи сигналов находят применение в качестве простейших моделей сигналов и помех неслучайные процессы, полностью известные в любой момент времени. Такие процессы называют детерминированными, их можно рассматривать как вырожденный класс случайных процессов, значения которых в любой момент времени известны с вероятностью, равной единице. Детерминированные процессы обычно используют как модели узкополосных сигналов-переносчиков и помех, случайные процессы — как модели полезных сигналов, шумоподобных, узкополосных и широкополосных сигналов-переносчиков и помех, модулированных сигналов.

Физическими характеристиками сигналов являются длительность T, ширина спектра $\Delta F_с$, динамический диапазон $Д = 10\lg(\mathscr{P}/N)$ и более общая характеристика — объем сигнала:

$$V_с = T\Delta F_с Д. \tag{1.1}$$

Длительность сигнала определяет время его существования, ширина спектра — диапазон частот, в котором сосредоточена основная энергия сигнала. Динамический диапазон характеризует отношение наибольшей мгновенной мощности сигнала ($\mathscr{P}_{макс} = \mathscr{P}$) к наименьшей ($\mathscr{P}_{мин}$), допустимое значение которой определяется мощностью помех $\mathscr{P}_п$ ($\mathscr{P}_{мин} \geqslant \mathscr{P}_п$). Эти характеристики сигналов полезны при определении требований, предъявляемых к каналам связи. Например, для неискаженной передачи сигналов емкость канала (см. § 1.3) должна быть не меньше объема сигнала.

Важной характеристикой сигналов является также база

$$\nu=2T\Delta F_{c}. \qquad (1.2)$$

Если $\nu\leqslant 1$, то сигналы называют узкополосными (простыми). Если $\nu\gg 1$, то — широкополосными (сложными).

Сущность большинства задач анализа реальных сигналов заключается в том, чтобы эти сигналы, которые в большинстве случаев являются сложными, представить в виде совокупности простых элементарных сигналов, удобном для последующего анализа их прохождения через те или иные цепи. Например, реальный сигнал может быть представлен в виде суммы ортогональных составляющих (элементарных сигналов)

$$s(t)=\sum_{k=0}^{\infty} a_k \psi_k(t), \quad t\in[t_1, t_2], \qquad (1.3)$$

бесчисленным количеством способов (см. § 2.1). Интервал $[t_1, t_2]$ показывает время действия сигнала. Так как система ортогональных функций $\{\psi_k(t)\}$, применяемая для разложения, заранее известна, то сигнал полностью определяется набором весовых коэффициентов a_k, $k=0, 1, 2, \ldots$, для этих функций. При приближенном представлении сигналов, что всегда имеет место в инженерной практике, набор чисел $\{a_k\}$ конечен. Такие наборы чисел называют *спектрами сигналов* [8].

Спектры, как известно из теории радиотехнических цепей [7], являются удобной аналитической формой представления сигналов в рамках линейной теории. Основная задача — правильный выбор системы ортогональных функций (базиса), удобной для последующего анализа прохождения сигнала через те или иные цепи и каналы связи. Таким образом, задачи анализа сигналов (анализа формы сигналов, их внутреннего строения, взаимосвязи элементов и т. п.) решают не отвлеченно, а, как правило, с точки зрения их прохождения через устройства передачи информации и способности передавать информацию и т. п.

Для детерминированных сигналов наибольшее распространение получили методы спектрального анализа, использующие преобразования Фурье. В этих методах в роли $\psi_k(t)$ выступают гармонические функции, а роль коэффициентов a_k играют амплитуды гармоник. Для случайных сигналов наибольшее распространение получили методы корреляционного и спектрального анализа, основанные на преобразовании Хинчина — Винера (см. § 2.3). Эти преобразования являются результатом распространения метода Фурье на случайные процессы. При разложении случайных процессов коэффициенты a_k являются случайными величинами, а оптимальные базисы определяются через корреляционные функции этих процессов.

Задачи синтеза сигналов могут быть двух типов: задачи структурного синтеза сигналов (задачи определения формы сигналов) и задачи параметрического синтеза (задачи определения парамет-

ров сигналов известной формы). Если в процессе синтеза стоит задача обеспечить экстремум того или иного функционала (или функции), который характеризует качество синтеза, то синтез называют оптимальным. Сущность задач синтеза сигналов рассмотрим на типичных примерах. Первая постановка задачи: определить такую оптимальную форму сигнала, при которой база сигнала будет минимальна. В этой задаче целевым функционалом структурного синтеза служит база сигнала, управляемой переменной — форма сигнала. Вторая постановка задачи: известны импульсная характеристика фильтра и энергия входного сигнала, требуется выбрать такую форму входного сигнала, чтобы энергия сигнала на выходе фильтра была максимальна. Третья постановка задачи: импульсная характеристика линейного фильтра и энергия входного сигнала заданы, требуется определить такую форму входного сигнала, при которой сигнал на выходе фильтра достигает максимума в заданный момент времени.

Рассмотренные постановки задач оптимального синтеза сигналов показывают, что эти задачи являются задачами на отыскание экстремума функции или функционала, в том числе и при наличии ограничений различного характера. Методы решения таких экстремальных задач разрабатываются в теории математического программирования. Задачи решают, как правило, с учетом того, в каких цепях будут циркулировать синтезируемые сигналы.

Рассмотренные примеры, конечно, не охватывают всех направлений синтеза сигналов. В теории информации и передачи сигналов изучают и задачи неоптимального синтеза, которые пока еще не решаются регулярными методами теории математического программирования. К задачам такого типа можно отнести, например, задачу получения требуемого числа высокостабильных колебаний, обладающих заданными свойствами и играющих роль сигналов-переносчиков, задачу получения требуемого числа ортогональных шумоподобных сигналов. По существу, это задачи синтеза технических устройств: датчиков опорных частот, генераторов шумоподобных сигналов и др.

Контрольные вопросы

1. Какими могут быть сообщения и сигналы?
2. Как рассматривают процессы формирования дискретных и непрерывных сообщений?
3. Какие математические модели сообщений, сигналов и помех используют?
4. Какими методами осуществляется построение математических моделей сообщений?
5. Какова цель математического описания сообщений, сигналов и помех?
6. Какие используют физические характеристики сигналов?
7. Что такое преобразование и линейность преобразования?
8. Что такое функция, оператор, функционал и функциональный оператор?
9. В чем сущность задач анализа сигналов?
10. В чем сущность задач оптимального синтеза сигналов? Укажите характерные примеры постановок таких задач.

1.2. МОДУЛЯЦИЯ КАК УПРАВЛЕНИЕ ИНФОРМАЦИОННЫМИ ПАРАМЕТРАМИ СИГНАЛОВ

Для передачи информации требуется, чтобы сигналы имели два вида параметров: параметры селекции (отбора) и информационные параметры. *Параметры селекции* позволяют выделить полезный сигнал из совокупности сигналов и помех. *Информационные параметры* служат для переноса информации — в изменении этих параметров отражаются сообщения. Управление информационным параметром переносчика в соответствии с законом изменения передаваемого сигнала (сообщения) в радиотехнике называют модуляцией.

Обозначим сигнал-переносчик через $x(t)$, передаваемый сигнал (сообщение) через $u(t)$. При модуляции выполняется преобразование двух сигналов $x(t)$ и $u(t)$ в один модулированный сигнал $s(t)$. Это преобразование обозначим оператором M_1:

$$s(t)=M_1[x(t),\ u(t)]. \qquad (1.4)$$

Для выделения переданного сигнала $u(t)$ из $s(t)$ необходимо выполнить демодуляцию $D_м$ — преобразование, обратное модуляции, что часто обозначают так: $D_м=M_1^{-1}$. Тогда

$$u(t)=D_м[s(t)]=M_1^{-1}[s(t)]. \qquad (1.5)$$

В зависимости от функциональной формы и числа параметров переносчика может быть большое число различных методов модуляции. Например, если в роли переносчика (несущей) выступает гармоническое колебание $x(a,\ \omega,\ \varphi;\ t)=a\cos(\omega t+\varphi)$, то может быть, как известно из курса радиотехнических цепей, три метода модуляции: амплитудная, частотная и фазовая. Применяют и комбинированную модуляцию, когда в соответствии с изменением передаваемого сигнала одновременно изменяются два независимых параметра переносчика. Примером служит однополосная модуляция, в результате которой получают амплитудно-частотно-модулированное колебание — однополосный сигнал. Во всех случаях один из параметров сигнала не должен изменяться, чтобы играть роль параметра селекции. Иначе, как уже отмечалось, модулированный сигнал невозможно будет выделить из множества других сигналов и помех. Важно отметить также и то, что один и тот же сигнал может переносить информацию об изменении n полезных сигналов, если он имеет n независимых информационных параметров.

Если под воздействием передаваемого сигнала информационный параметр переносчика изменяется непрерывно, все возможные виды модуляции являются непрерывными. К ним относят, например, амплитудную, фазовую и частотную модуляцию гармонического колебания. Если в роли переносчика используют периодическую последовательность импульсов, то модуляцию называют импульсной. Например, к импульсным видам модуляции относят

амплитудно-импульсную, частотно-импульсную и широтно-импульсную модуляции. Если при модуляции информационный параметр принимает счетное число значений, такую модуляцию называют дискретной. К дискретным видам модуляции относятся амплитудная, частотная и фазовая манипуляции [3]. Если эти счетные значения пронумеровывают и в виде цифр передают по линии связи, то говорят о цифровой модуляции. Примерами таких видов модуляции служат импульсно-кодовая модуляция, дельта-модуляция [1—3].

Основной задачей управления информационными параметрами сигналов является разработка методов и математических моделей, которые позволяют выполнить анализ и синтез операторов модуляции и демодуляции, желательно совместно с анализом сигналов $u(t)$ и $s(t)$ и синтезом сигналов $x(t)$.

Эти операторы реализуют с помощью функциональных преобразователей, которые называют соответственно модуляторами и демодуляторами. С инженерной точки зрения задачи анализа и синтеза таких операторов сводятся, по существу, к задачам анализа и синтеза модуляторов и демодуляторов, а также и генераторов сигналов-переносчиков. Так как модуляция и демодуляция образуют связанную пару операторов, целесообразно анализировать и синтезировать модемы (устройства модуляции и демодуляции) совместно с генераторами сигналов.

Контрольные вопросы

1. Какие виды параметров должны иметь сигналы, чтобы они могли переносить информацию?
2. Как записывают в операторной форме операции модуляцию и демодуляцию?
3. Можно ли с помощью одного сигнала передавать информацию от нескольких источников?
4. Укажите основные виды модуляции гармонического колебания и импульсной последовательности?
5. К чему сводятся задачи анализа и синтеза операторов модуляции и демодуляции?

1.3. КАНАЛЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

В общем случае *под каналом передачи информации* понимают всю совокупность технических средств, обеспечивающих передачу электрических сигналов от источника сообщений к потребителю. При рассмотрении каналов линию связи чаще всего полагают заданной (считается, что все необходимые характеристики линии связи известны) и все задачи анализа и синтеза каналов передачи информации сводятся к анализу и синтезу операторов преобразования сигналов в передатчике, приемнике и других устройствах.

Каналы передачи информации классифицируют по различным признакам: по назначению, по характеру линий связи, по диапазону частот, по характеру сигналов на входе и выходе каналов и т. п. По назначению каналы делятся на телефонные, телеграф-

ные, телевизионные, фототелеграфные, звукового вещания, телеметрические, передачи данных и др. В зависимости от того, распространяются ли сигналы в свободном пространстве или по направляющим линиям, различают каналы радиосвязи и каналы проводной связи: воздушные, кабельные, волноводные, световодные и др. По воздушным проводным линиям связи передают сигналы в диапазоне 0—160 кГц. На более высоких частотах возрастает влияние помех, резко увеличивается затухание сигналов, сказывается влияние радиовещательных станций длинноволнового диапазона. Существенный недостаток воздушных проводных линий связи — большая зависимость их характеристик от атмосферных условий. Значительно лучшими характеристиками и большей устойчивостью в работе обладают кабельные линии связи. Они являются основой сетей магистральной дальней связи, по ним передают сигналы в диапазоне частот от 600 кГц до 60 МГц. С дальнейшим увеличением частоты затухание сигналов резко возрастает.

В настоящее время ведутся интенсивные теоретические и экспериментальные работы по исследованию металлических волноводов. Полученные результаты позволяют надеяться, что эти линии связи будут широко использоваться для передачи сигналов в диапазоне 35—80 ГГц (длина волны 8,6—3,75 мм). Перспективен круглый волновод с внутренним диаметром 6 см, по которому можно организовать более 200 000 стандартных телефонных каналов (каналов тональной частоты с эффективно используемой полосой частот от 300 до 3400 Гц) или около 200 телевизионных каналов [8]. Экономические расчеты показывают, что при организации телефонных каналов до 30 000 еще целесообразно применять коаксиальный кабель, свыше 30 000 каналов — волновод. Еще большее число стандартных каналов можно организовать, используя оптические системы связи, в которых применяют сигналы в полосе частот 600—900 ТГц (0,5—0,3 мкм). Используя закрытые направляющие системы, которые получили название световодов, можно осуществить устойчивую связь на большие расстояния. Большой практический интерес представляют диэлектрические гибкие волоконные световоды.

Наряду с проводными линиями связи широко используют радиолинии различных диапазонов. Эти линии во многих случаях более экономичны, позволяют быстро организовать сверхдальнюю (глобальную) связь без промежуточных станций. Кроме того, и это очень важно,— эти линии являются единственным средством связи с подвижными объектами (воздушными судами, космическими кораблями, морскими судами, включая и подводные лодки, автомобилями и пр.).

Наибольшее распространение для передачи многоканальных сообщений получили наземные радиорелейные линии, работающие в метровом, дециметровом и сантиметровом диапазонах волн на частотах от 60 МГц до 15 ГГц. На этих частотах обеспечивается широкая полоса тракта передачи, необходимая для многоканальной телефонной и телевизионной связи, мал уровень атмос-

ферных и промышленных помех. Все это обеспечивает высокую помехоустойчивость передачи информации.

Разновидностью радиорелейных линий являются тропосферные линии, в которых принимаются сигналы, отраженные от неоднородностей тропосферы. Использование дальнего тропосферного распространения радиоволн позволяет создать линии дальней радиосвязи с расстояниями между ретрансляционными станциями в несколько сотен километров. Эти линии работают чаще всего в диапазоне частот от 0,5 до 6 ГГц.

Перспективны спутниковые линии связи. По принципу работы они представляют разновидность радиорелейных линий, ретрансляторы которых находятся на искусственных спутниках Земли. Существенным преимуществом спутниковых линий является большая дальность связи, которая при одном спутнике (ретрансляторе) составляет около 10 000 км. При использовании системы спутников можно организовать глобальную связь — между любыми пунктами Земли. Спутниковые линии связи работают в диапазоне частот 4—6 ГГц. В настоящее время отведено шесть новых частотных диапазонов от 11 до 250 ГГц, освоение которых позволит существенно повысить качественные показатели спутниковой связи. Спутниковые системы связи, особенно с цифровыми методами передачи сигналов, перспективны и в гражданской авиации, особенно с выходом на воздушные трассы сверхзвуковых пассажирских судов.

Как видим, для современных методов и средств передачи информации характерен переход на все более высокие частоты. Это обусловлено следующими основными причинами: применение высоких частот позволяет получить остронаправленное излучение при малых размерах антенн; в высокочастотных диапазонах меньшее влияние оказывают атмосферные и промышленные помехи; чем выше несущая частота, тем большее число каналов можно организовать без взаимных помех; только в высокочастотных диапазонах, начиная с метрового, можно организовать большое число широкополосных каналов, таких, например, как каналы видеотелефонной связи и телевизионные каналы.

Одной из основных задач анализа каналов передачи информации является анализ искажений передаваемых по ним сигналов. Более всего на качестве передачи информации сказываются искажения формы сигналов, определяемые реальными амплитудными и частотными характеристиками каналов, а также многолучевым распространением радиоволн. Математические модели для полного анализа искажений в реальных каналах достаточно сложны.

Для общего приближенного анализа искажений канал рассматривают как эквивалентный четырехполюсник, работа которого описывается определенным оператором L. Если входной сигнал $s_1(t)$, то сигнал на выходе канала $s_2(t)=L[s_1(t)]$. Для более детального анализа искажений канал связи представляют как последовательное соединение линейного, в общем случае инерционного, и нелинейного неинерционного четырехполюсников, в кото-

рых и происходят линейные и нелинейные искажения сигналов. Если в состав канала входят модулятор и демодулятор или учитываются замирания сигналов, то к этим четырехполюсникам последовательно включают еще и четырехполюсник с переменными параметрами. Если требуется анализировать работу отдельных устройств канала, число последовательно соединенных четырехполюсников увеличивают. Например, для анализа радиоканала передачи дискретных сообщений часто используют структурную схему, представленную на рис. 1.1. Канал рассматривается как последо-

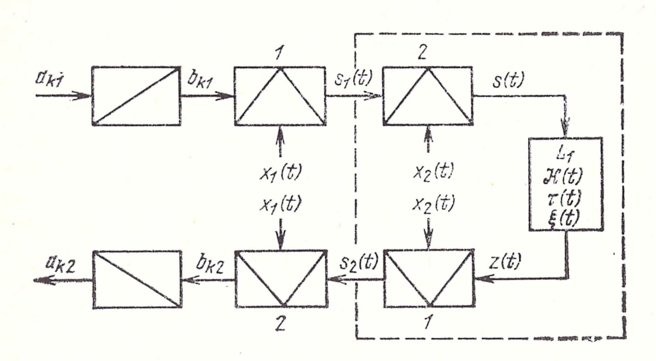


Рис. 1.1. Структурная схема канала передачи информации

вательное соединение кодера, первого и второго модулятора, линии связи, первого и второго демодулятора и декодера. Использование модели канала в виде эквивалентного четырехполюсника (или последовательного соединения четырехполюсников) позволяет решить ряд задач анализа и синтеза каналов методами теории радиотехнических цепей [7] и статистической радиотехники [15].

По характеру сигналов на входе и выходе каналов различают дискретные, непрерывные, дискретно-непрерывные и непрерывно-дискретные каналы. *Дискретным каналом*, например в схеме рис. 1.1, является канал, рассматриваемый от входа кодера до выхода декодера или от входа первого модулятора до выхода второго демодулятора. Если рассматривать совокупность технических средств от выхода первого или второго модулятора до входа первого или второго демодулятора, то в любой системе передачи информации эта совокупность образует *непрерывный канал*. Примером *дискретно-непрерывного канала* служит совокупность технических средств от входа первого модулятора до входа первого или второго демодулятора. *Непрерывно-дискретный канал* образуется, если анализировать в схеме рис. 1.1 прохождение сигналов от выхода первого или второго модулятора до выхода второго демодулятора или в общем случае до выхода декодера. Таким образом, на основе непрерывного канала (на рис. 1.1 он показан штриховой

линией) образуются каналы всех других видов. Поэтому изучению непрерывных каналов уделяют большое внимание.

Математические модели для исследования каналов строят с учетом рассмотренной классификации. По существу разработка модели сводится к определению структуры и параметров эквивалентного оператора преобразования сигнала в канале. В зависимости от типа этого оператора различают каналы: линейные, нелинейные, инерционные, безынерционные, стационарные, нестационарные, детерминированные, вероятностные и др. Наиболее изучены линейные инерционные каналы с постоянными параметрами.

Если передаваемый сигнал $s(t)$ рассматривается как случайный процесс, что значительно приближает модели сигналов к реальным, то при анализе прохождения сигнала $s(t)$ через канал необходимо анализировать прохождение случайных процессов через четырехполюсники, описываемые различными операторами. Такие задачи решают в статистической радиотехнике [15], результаты решения этих задач находят непосредственное применение при анализе каналов передачи информации.

Как и для сигналов, для каналов удобно использовать такие физические характеристики, как время занятости канала $T_к$, полоса $\Delta F_к$ пропускания (прозрачности) канала, диапазон $Д_к$ допустимых уровней сигналов при передаче по каналу, база канала $\nu_к = 2T_к\Delta F_к$, емкость канала

$$V_к = T_к \Delta F_к Д_к. \quad (1.6)$$

Контрольные вопросы

1. Как определяют канал передачи информации?
2. Укажите признаки классификации каналов.
3. Назовите основные типы проводных линий связи.
4. Назовите основные типы радиолиний связи.
5. Чем обусловлена тенденция к переходу на более высокие частоты связи?
6. Как математически описывают работу каналов?
7. Нарисуйте и поясните структурную схему радиоканала для передачи дискретных сообщений.
8. Какие методы используют для анализа работы каналов?
9. Как различают каналы по характеру сигналов на их входе и выходе?
10. Назовите основные физические характеристики каналов.

1.4. ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ИСТОЧНИКОВ СООБЩЕНИЙ И КАНАЛОВ

Основными информационными характеристиками являются количество информации в сообщениях, избыточность сообщений, энтропия и производительность источника сообщений, скорость передачи информации и пропускная способность канала. Одна из важных задач теории информации и передачи сигналов — разработка методов расчета этих характеристик. Рассмотрим указанные характеристики для случая передачи дискретных сообщений.

Обозначим объем алфавита **A** источника дискретных сообщений через m. Предположим, что каждое сообщение включает n

символов. Покажем, как определяется количество информации в сообщениях такого источника.

При выборе способа измерения количества информации учитывают следующее: сообщения большей длины содержат, как правило, большее количество информации; большее количество информации содержится и в тех сообщениях, которые образуются из символов алфавита, имеющего больший объем; формирование сообщений носит случайный характер, символы сообщений могут появляться с различными вероятностями и могут быть статистически зависимыми; более полезные для получателя информации сообщения содержат большее количество информации.

Общее число различных сообщений длиной n $N_0=m^n$. Это число можно было бы использовать как некоторую информационную характеристику источника сообщений, но она неудобна из-за степенного характера зависимости N_0 от n. Р. Хартли предложил в 1928 г. логарифмическую меру количества информации

$$I=\log N_0=n \log m. \qquad (1.7)$$

Формула (1.7) Хартли не отражает случайного характера формирования сообщений. Чтобы устранить этот недостаток, необходимо связать количество информации в сообщениях с вероятностями появления символов. Эту задачу решил в 1946 году К. Шеннон.

Если вероятности появления всех символов алфавита одинаковы, то количество информации, которое переносит один символ, $I_1=\log m$. Вероятность появления символов $p=1/m$, следовательно, $m=1/p$. Подставив это значение m в формулу для I_1, получим $I_1=-\log p$. Полученное соотношение уже связывает количество информации, которое переносит один символ, и вероятность появления этого символа. В реальных сообщениях символы появляются с различными вероятностями, символ a_i появляется с вероятностью $p(a_i)$, $a_i \in \mathbf{A}$. Количество информации, которое переносит этот символ, $I_i=-\log p(a_i)$. Среднее количество информации $H(\mathbf{A})$, которое приходится на один символ источника сообщений, получим, применив операцию усреднения по всему объему алфавита

$$H(\mathbf{A})=-\sum_{i=1}^{m} p(a_i) \log p(a_i). \qquad (1.8)$$

Выражение (1.8) известно как формула Шеннона для *энтропии источника дискретных сообщений*. Энтропия рассматривается как мера неопределенности в поведении источника.

Как непрерывная функция вероятностей появления символов энтропия обладает следующими свойствами: энтропия источника дискретных сообщений есть величина вещественная, ограниченная и положительная; энтропия равна нулю, если с вероятностью единица всегда выбирается один и тот же символ (в этом случае неопределенность в поведении источника сообщений отсутствует);

энтропия максимальна, если все символы источника появляются независимо и с одинаковой вероятностью, и равна

$$H_{max} = -mp\log p = -m\frac{1}{m}\log\frac{1}{m} = \log m. \qquad (1.9)$$

Сравнение формул (1.7), (1.9) показывает, что способ измерения количества информации, предложенный К. Шенноном, является обобщением способа Хартли на случай появления неравновероятных независимых символов.

Чтобы учесть статистические (корреляционные) связи между символами сообщений, вводят понятие условной энтропии

$$H(A'|A) = -\sum_{i=1}^{m} p(a_i)\sum_{j=1}^{m} p(a'_j|a_i)\log p(a'_j|a_i), \qquad (1.10)$$

где $p(a'_j|a_i)$ — вероятность появления a'_j при условии, что перед ним появился a_i. Условная энтропия (1.10) — это среднее количество информации, которое переносит один символ сообщений при условии, что существуют корреляционные связи между двумя соседними символами в сообщениях.

Из-за корреляционных связей между символами и неравновероятного их появления в реальных сообщениях падает количество информации, которое переносит один символ. Количественно эти потери информации характеризуют *коэффициентом избыточности*

$$r = (H_1 - H_2)/H_1 = 1 - H_2/\log m, \qquad (1.11)$$

где H_1 — максимальное количество информации, которое может переносить один символ; H_2 — количество информации, которое переносит один символ в реальных сообщениях. Для европейских языков избыточность сообщений $r \geqslant 0{,}5$.

Как единица измерения количества информации в сообщениях наибольшее распространение получил один *бит* — максимальное количество информации, которое переносит один символ источника дискретных сообщений в том случае, когда алфавит источника включает два символа и используется двоичная система логарифмов. Слово бит образовалось путем сокращения английского названия binary digit (двоичная цифра).

Среднее количество информации, выдаваемой источником в единицу времени, называют производительностью источника.

$$H' = H/t\,[\text{бит/с}]. \qquad (1.12)$$

Для каналов передачи информации вводят аналогичную характеристику — скорость передачи информации по каналу (среднее количество информации, которое можно передать по каналу в единицу времени)

$$R = WH\,[\text{бит/с}], \qquad (1.13)$$

где W — скорость передачи электрических кодовых сигналов; H — среднее количество информации, которое несет один кодовый сигнал.

Максимальное значение скорости передачи информации по каналу называют пропускной способностью канала. Если помехи и стирания в канале отсутствуют [см. § 4.3], пропускная способность дискретного канала

$$C = \max WH = W \max H = W \log m. \qquad (1.14)$$

Согласование производительности источника и пропускной способности канала — одна из наиболее важных задач кодирования.

С помощью информационных характеристик анализируют и сравнивают различные источники сообщений и каналы, выполняют согласование источников и каналов, определяют влияние помех на передачу информации, создают коды, обеспечивающие максимальную скорость или помехоустойчивость передачи информации и решают другие практически важные задачи. Разработке способов оценки (измерения) с точки зрения качества и ценности информации уделяется значительное внимание.

Контрольные вопросы

1. Назовите основные информационные характеристики источников сообщений и каналов.
2. Что необходимо учитывать при выборе способа измерения количества информации?
3. Укажите достоинства и недостатки способов, предложенных Р. Хартли и К. Шенноном.
4. Какими свойствами обладает энтропия источника дискретных сообщений?
5. Почему вводится понятие условной энтропии, как определяют избыточность сообщений?
6. Дайте определения производительности источника, скорости передачи информации и пропускной способности канала.

1.5. ПОМЕХОУСТОЙЧИВОСТЬ ПЕРЕДАЧИ ИНФОРМАЦИИ

Из-за влияния помех полученные сообщения отличаются от оригиналов. *Помехоустойчивостью* системы называют ее способность передавать информацию с заданной верностью при воздействии помех. Количественно помехоустойчивость характеризуют степенью соответствия принятого сообщения переданному. Эту величину называют *критерием верности* (достоверности) передачи информации. При передаче дискретных сообщений влияние помех проявляется в том, что вместо того или иного переданного символа принимается другой. Такое случайное событие называют ошибкой. Простейшим критерием верности при передаче дискретных сообщений является вероятность появления ошибки при передаче одного символа или одного бита информации.

При передаче непрерывных сообщений степень соответствия переданного и принятого сигналов характеризует случайная величина отклонения принятого сигнала $z(t)$ от переданного $s(t)$. Мерой отклонения обычно служит расстояние ε между $z(t)$ и $s(t)$, выбранное в функциональном пространстве сигналов [см. § 2.10]. Критерием верности в этом случае является вероятность p_0 того,

что наблюдаемое отклонение будет меньше некоторого заданного ε_0: $p_0 = p(\varepsilon < \varepsilon_0)$.

Основы теории помехоустойчивости были заложены В. А. Котельниковым в фундаментальной работе «Теория потенциальной помехоустойчивости» (1946 г.). *Потенциальной помехоустойчивостью* В. А. Котельников назвал предельно достижимую помехоустойчивость передачи информации при заданной помехе. Дальнейшая разработка методов анализа и синтеза помехоустойчивых способов передачи информации — одна из важнейших проблем теории информации и передачи сигналов. Основными задачами этой проблемы являются: выбор и обоснование критериев верности для различных условий передачи информации; анализ помехоустойчивости методов и алгоритмов передачи информации; синтез помехоустойчивых методов и алгоритмов передачи информации; техническая реализация оптимальных методов и алгоритмов передачи информации (разработка технических средств для оптимальной передачи информации) и др.

В настоящее время наиболее изучены относительно простые задачи обработки сигналов в приемных устройствах при наличии помех в линии связи. Их обычно решают в предположении, что способ передачи и характеристики помех в каналах априорно известны. Менее изучены задачи совместного анализа и оптимального синтеза способов передачи и способов приема. Это обусловлено значительной сложностью таких задач.

Контрольные вопросы

1. Как определяют помехоустойчивость передачи информации?
2. Что такое критерий верности?
3. Что такое ошибка при передаче дискретных сообщений?
4. Как выбирают критерии верности при передаче дискретных и непрерывных сообщений?
5. Что понимают под потенциальной помехоустойчивостью?
6. Сформулируйте основные задачи теории помехоустойчивости.

1.6. КОДИРОВАНИЕ

Различают *эффективное* (статистическое) и *корректирующее* (помехоустойчивое) *кодирование*. Целью эффективного кодирования является повышение скорости передачи информации и приближение ее к пропускной способности каналов. Теоретической основой построения эффективных кодов служит *теорема Шеннона*, в которой утверждается, что для канала без помех всегда можно создать систему эффективного кодирования дискретных сообщений, у которой среднее количество двоичных кодовых сигналов на один символ сообщения будет приближаться как угодно близко к энтропии источника сообщений. Целью помехоустойчивого кодирования является повышение верности передачи информации путем обнаружения и исправления ошибок. Теоретической основой помехоустойчивого кодирования является *теорема Шеннона*, в которой

утверждается, что и для канала с помехами всегда можно найти такую систему кодирования, при которой сообщения будут переданы со сколь угодно большой степенью верности, если только производительность источника не превышает пропускной способности канала.

При кодировании каждый символ дискретного сообщения пронумеровывается и передача сообщений сводится к передаче последовательностей чисел, записанных в той или иной системе счисления. Например, для передачи букв русского алфавита необходимо передавать числа от 1 до 32 (твердый знак не передают). Обозначим основание системы счисления через γ (например, в десятичной системе $\gamma=10$, в двоичной $\gamma=2$). Любое n-разрядное число χ можно записать в виде полинома

$$\chi=\sum_{i=1}^{n} a_i\gamma^i, \tag{1.15}$$

где a_i — целые числа, удовлетворяющие условию $0\leqslant a_i\leqslant\gamma-1$. Например, число 13 в десятичной и двоичной системе записывается так: $(13)_{\gamma=10}=1\cdot10^1+3\cdot10^0$; $(13)_{\gamma=2}=1\cdot2^3+1\cdot2^2+0\cdot2^1+1\cdot2^0=$ $=1101$.

Когда последовательность элементов дискретного сообщения преобразована в последовательность двоичных чисел, для их передачи по каналу связи необходимы всего лишь два элементарных кодовых сигнала: 0, 1. Роль таких сигналов могут выполнять посылки постоянного тока разной полярности, колебания с различными частотами и др. При передаче двоичных сигналов требуется всего лишь пара проводов. Другим важным преимуществом двоичного кодирования является простота арифметических операций в двоичной системе счисления, которые легко реализуются с помощью простейших логических схем вычислительной техники. Сложение выполняется схемой ИЛИ, а умножение — схемой И. Более сложные операции реализуются различными сочетаниями схем И и ИЛИ.

Таким образом, *кодирование* — это процесс преобразования элементов дискретного сообщения в соответствующие числа, выраженные кодовыми символами. *Кодом* называют полную совокупность условных символов, которую применяют для кодирования сообщений. Число различных символов в коде называют *основанием кода* $m_к$. *Кодовая комбинация* — это последовательность кодовых символов, соответствующая одному элементу дискретного сообщения (число, записанное в системе счисления с основанием $m_к$). *Значность кода* n — это число символов в кодовой комбинации. *Оператор кодирования* показывает, какую кодовую комбинацию присваивают каждому элементу сообщения.

Если $m_к=2$, код называют двоичным (бинарным), при $m_к>$ >2 — многопозиционным. Если все кодовые комбинации содержат одинаковое число символов, код называют равномерным, в других

случаях — неравномерным. Для равномерного кода общее число различных кодовых комбинаций $N = m^{n}_{\text{к}}$. Например, для равномерного кода Бодо $m_{\text{к}} = 2$, $n = 5$, $N = 32$. Этого числа комбинаций достаточно для кодирования всех букв латинского алфавита. Код Морзе служит примером неравномерного кода. В нем наиболее часто встречающимся буквам присваиваются наиболее короткие кодовые комбинации. Это сделано для увеличения скорости передачи информации.

Рассмотрим сущность идей обнаружения и исправления ошибок корректирующими кодами. Идея обнаружения ошибок заключается в следующем. Для передачи используют не все N кодовых комбинаций, а только часть из них N_0, которые называют разрешенными. Оставшиеся $\Delta N = N - N_0$ комбинаций называют запрещенными. Ошибки обнаруживают тогда, когда на приемном конце получают запрещенную комбинацию. Всякий код, у которого $\Delta N > 0$, способен обнаруживать ошибки в ΔN случаях из N. Доля обнаруживаемых ошибок

$$\Delta N / N = 1 - N_0 / N. \tag{1.16}$$

Если $\Delta N = 0$, т. е. $N = N_0$, код не способен обнаруживать ошибки и его называют примитивным (безызбыточным).

Избыточность корректирующего кода

$$r_{\text{к}} = 1 - (\log N_0) / n \log m_{\text{к}}. \tag{1.17}$$

Из формул (1.16), (1.17) следует, что корректирующий код обнаруживает ошибки за счет введения избыточности. Доля обнаруживаемых ошибок растет по мере увеличения избыточности.

Исправление ошибок корректирующими кодами основано на двух операциях: определении расстояния между кодовыми комбинациями и отыскании минимального расстояния. *Расстоянием* d_{ij} между кодовыми комбинациями K_i и K_j называют суммарный результат сложения по модулю $m_{\text{к}}$ одноименных разрядов кодовых комбинаций

$$d_{ij} = \sum_{k=1}^{n} \kappa_{ik} \oplus \kappa_{jk}, \tag{1.18}$$

где κ_{ik} и κ_{jk} — k-й разряд кодовых комбинаций K_i и K_j; символ $\oplus$ обозначает сложение по модулю m_k; n — значность кода. Сущность суммирования по модулю m_k заключается в том, что результат суммирования равен модулю суммы разрядов, если этот модуль меньше m_k. Если модуль суммы больше m_k, то результат получают вычитанием величины m_k из суммы.

Аналитическая запись сложения по модулю m_k имеет вид

$$\kappa_{ik} \oplus \kappa_{jk} = \begin{cases} \kappa_{ik} \pm \kappa_{jk}, & \kappa_{ik} \pm \kappa_{jk} < m_k; \\ \kappa_{ik} \pm \kappa_{jk} - m_k, & \kappa_{ik} \pm \kappa_{jk} \geqslant m_k. \end{cases} \tag{1.19}$$

На рис. 1.2 показано условное обозначение логического оператора сложения по модулю. Таким образом, расстояние получают в результате поразрядного суммирования по модулю и обычного суммирования по формуле (1.18). Для равномерного двоичного кода кодовое расстояние — это число символов, на которое отличается одна комбинация от другой. Например, если $K_i=10111$, $K_j=$ $=01010$, то $d_{ij}=4$.

Идея исправления ошибок заключается в том, что, обнаружив ошибку, определяют расстояние от полученной запрещенной комбинации K_i до всех разрешенных K_j, $j=\overline{1, N_0}$, и в качестве переданной считают ту из разрешенных комбинаций, до которой расстояние меньше всего. Например, если $\min d_{ij}=$ $=d_{i5}$, $j=\overline{1, N_0}$, то полагают, что была передана комбинация K_5.

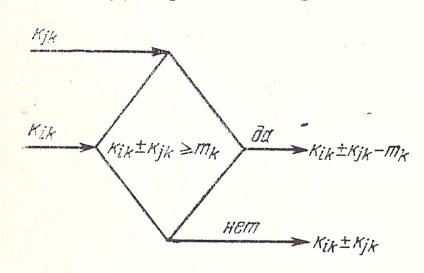


Рис. 1.2. Логический оператор сложения по модулю

Задачи анализа и синтеза кодов, устройств кодирования и декодирования решают на основе сочетания методов теории информации, высшей алгебры и теории конечных автоматов [18]. Методы теории информации позволяют строить общие алгоритмы анализа и синтеза кодов, методы высшей алгебры (теория групп, теория колец, теория полей) позволяют синтезировать и анализировать математические модели кодов, методы теории конечных автоматов — создавать конкретные структурные схемы кодеров и декодеров.

Контрольные вопросы

1. Укажите цель эффективного и корректирующего кодирования.
2. Что служит теоретической основой эффективного и корректирующего кодирования?
3. Как записать n-разрядное число в виде полинома?
4. Дайте определения кодирования, кода, кодовой комбинации, значности кода, оператора кодирования.
5. В чем сущность обнаружения ошибок?
6. Как определяют расстояние между кодовыми комбинациями?
7. В чем сущность исправления ошибок?
8. Какие методы применяют для анализа и синтеза кодов, устройств кодирования и декодирования?

1.7. УПЛОТНЕНИЕ ЛИНИЙ СВЯЗИ. ИНФОРМАЦИОННЫЕ ПОТОКИ В СЕТЯХ

Методы уплотнения линий связи — это, по существу, методы построения многоканальных систем передачи информации. Задачи анализа многоканальных систем включают уже рассмотренные задачи анализа одноканальных систем и, кроме того, задачи исследования влияния структуры и параметров операторов уплотнения и разделения сигналов на характеристики качества передачи информации. Задачи оптимального синтеза многоканальных сис-

тем включают задачи синтеза одноканальных систем и, кроме того, задачи определения оптимальных структур и параметров операторов уплотнения и разделения, которые доставляют экстремум выбранному функционалу качества при выполнении ряда технико-экономических ограничений. Отсюда следует, что задачи анализа и синтеза многоканальных систем, задачи разработки эффективных методов уплотнения линий связи являются более сложными. Эти задачи рассматривают в курсах «Системы связи» и «Связное оборудование».

Основная проблема теории многоканальных систем — создание методов эффективного разделения канальных сигналов. Эти методы разрабатывают в теории селекции сигналов. В роли канальных сигналов-переносчиков чаще всего используют ортогональные сигналы, гармонические колебания, колебания более сложных функциональных форм, шумоподобные сигналы, различные импульсные последовательности и др. В роли параметров селекции сигналов обычно используют частоту, фазу, форму, время поступления и т. п. Многоканальные системы уплотнения линий связи являются системами первичного уплотнения, так как каналы этих систем могут еще раз уплотняться. Уплотнение каналов многоканальных систем получило название вторичного уплотнения.

Каналы связи сетей, имеющие устройства уплотнения, включают также устройства контроля качества передачи информации, обнаружения и исправления ошибок, усиления и ретрансляции и др. Канал связи и устройства передачи и приема сигналов образуют *тракт передачи информации. Абонентскими пунктами* называют источники и потребители информации с соответствующей аппаратурой передачи и приема информации по каналам связи. На одном и том же абонентском пункте может находиться как аппаратура передачи, так и аппаратура приема информации. Линии связи и абонентские пункты рассматривают *как устройства передачи информации.*

Соединение (коммутацию) между собой отдельных каналов связи выполняют *устройства распределения информации.* Два или более тракта передачи информации, скоммутированных последовательно с помощью устройств распределения информации, образуют соединительный тракт передачи информации. Если между двумя абонентскими пунктами образован соединительный тракт, то говорят, что между ними скоммутирован *составной канал связи.* Если коммутируют многоканальные линии связи, то говорят о широкополосной коммутации линий (каналов). На рис. 1.3 показаны основные элементы сетей: линии связи, абонентские линии, абонентские пункты, каналообразующая аппаратура, устройства распределения и накопления информации. *Абонентской линией* называют одноканальную линию связи, которая непосредственно соединяет абонентский пункт с устройствами распределения и накопления информации.

Линии связи развиваемой в настоящее время единой автоматизированной сети связи страны (ЕАСС) образуют *первичную сеть.*

На ее основе с использованием устройств распределения и накопления информации создают вторичные сети различного назначения. Вторичная сеть связи — это совокупность трех основных элементов: сети каналов связи, системы коммутации и системы управления.

Система коммутации предназначена для образования составных каналов связи или для коммутации сообщений. Система управления обеспечивает управление информационными потоками в сети с учетом реально сложившейся ситуации. Ситуацию на сети

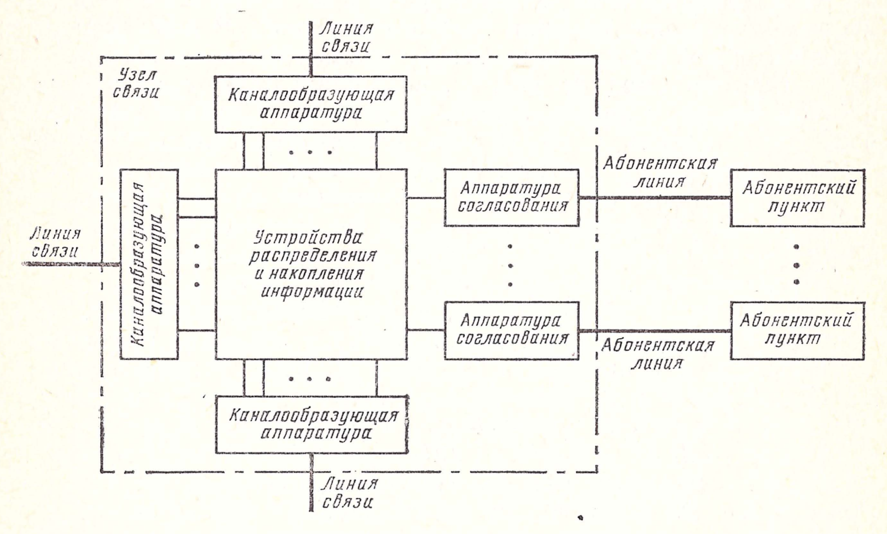


Рис. 1.3. Основные элементы сети

определяют изменения интенсивностей информационных потоков (интенсивность потока сообщений — это количество сообщений, вырабатываемых в единицу времени источником информации), исправность и качество работы каналов и узлов связи, приоритет поступающих сообщений, использование отдельных каналов и узлов связи, изменение уровней помех и т. п.

Эффективность использования каналов и узлов связи характеризуют коэффициентом использования, который определяется как отношение интенсивности входящего информационного потока к емкости канала или узла. Под *емкостью канала или узла* понимают ту интенсивность потока сообщений, которая может быть передана по этому каналу при заданном качестве передачи информации.

Вторичные сети связи классифицируют по многим признакам: функциональному назначению, принципу построения и количеству уровней иерархии структуры, способу коммутации, способу и дисциплине обслуживания абонентов, скорости передачи информации, способу управления информационными потоками, характеру

изменения во времени структуры и параметров алгоритмов управления и др. По функциональному признаку сети связи гражданской авиации разделяют на авиационные (наземные и воздушные, предназначенные для непосредственного управления воздушным движением), аэропортовые (предназначенные для управления производственно-технологической деятельностью служб аэропортов), коммерческие (предназначенные для управления производственной деятельностью всех звеньев служб перевозок и коммерческой эксплуатации), международные (наземные и воздушные, предназначенные для обеспечения взаимодействия диспетчерских служб СССР и других стран при полетах самолетов Аэрофлота и иностранных авиакомпаний по международным воздушным линиям), командно-служебные (предназначенные для обеспечения руководства предприятиями, организациями и учреждениями гражданской авиации).

Основные принципы построения структуры сетей следующие:

— радиально-узловой принцип, когда группа абонентских пунктов обслуживается одним общим узлом связи, который через центральный узел связи осуществляет соединения с другими узлами и, следовательно, с абонентами других групп; узлы связи устанавливают соединение только через центральный узел;

— принцип соединения узлов сети «каждый с каждым»;

— принцип использования решетчатой структуры (принцип построения распределенных сетей), когда все узлы связи, через которые проходят большие информационные потоки (узлы первого класса), соединены по принципу «каждый с каждым», а абонентские пункты или узлы с меньшими информационными потоками (узлы второго класса) соединены с ближайшими узлами первого класса;

— иерархический принцип, когда образуется ряд подсетей, образующих по определенному признаку иерархическую структуру.

По способу коммутации различают сети с коммутацией каналов, сообщений, а также с коммутацией каналов и сообщений. *Коммутацией каналов* называют совокупность операций, состоящих в соединении нескольких каналов связи для получения составного (сквозного) канала связи, используемого для передачи информации от источника к потребителю. Составной канал образуется только на время передачи информации, например на время телефонного разговора. После использования составного канала все соединения ликвидируются. Если один из транзитных каналов составного канала занят или если занят абонент, передача информации не состоится. Следовательно, в сетях с коммутацией каналов имеет место способ обслуживания с отказами в соединении. Такие сети часто называют системами обслуживания с потерями.

Коммутация сообщений — это совокупность операций, состоящих в приеме сообщений, их накоплении и последующей передаче в соответствии с содержащимися в них адресами. В адресе указан номер потребителя информации, категория сообщения (приоритет), допустимое число *переприемных пунктов* и т. п. В сетях

с коммутацией сообщений узлы связи содержат устройства распределения и накопления информации. Каждое сообщение поэтапно передается и запоминается в узлах связи. Занятость каналов не играет роли, так как передача сообщения к потребителю выполняется по мере освобождения каналов требуемого направления. Следовательно, по способу обслуживания абонентов сети с коммутацией сообщений являются системами обслуживания с ожиданием. В узлах связи сетей с коммутацией сообщений передача информации от входного канала к выходному осуществляется через запоминающее устройство системы коммутации. Такие узлы связи получили название центров коммутации сообщений.

В последнее время все большее внимание уделяют развитию сетей с *коммутацией каналов и сообщений*. По способу обслуживания эти сети являются системами обслуживания с ограниченным ожиданием — с ограниченными потерями. В таких сетях крупные узлы связи выполняют как центры коммутации сообщений, а мелкие периферийные узлы как центры коммутации каналов. Развитие элементной базы вычислительной техники, унификация и стандартизация функциональных узлов систем коммутации и управления потоками в сетях на основе функциональных узлов цифровой вычислительной техники позволяют существенно снизить стоимость оборудования центров коммутации линий и сообщений, что способствует успешному развитию таких сетей.

По скорости передачи информации различают *низкоскоростные сети* (скорости передачи информации —50, 100 и 200 бит/с), *среднескоростные* (скорости —600, 1200, 2400 и 4800 бит/с), *высокоскоростные* (скорости — более 4800 бит/с) и *сети смешанного типа*, скорости передачи информации в различных ветвях которых изменяются в широких пределах.

По способу управления информационными потоками сети делят на сети с централизованным, децентрализованным и локализованным управлением. При централизованном управлении все процедуры выбора направления движения информации осуществляются в едином центральном узле связи. К сетям с централизованным управлением относятся, например, диспетчерские сети управления воздушным движением. При децентрализованном управлении функции управления распределены между различными узлами сети. Локализованное (зонное) управление — это промежуточный тип управления, функции управления распределены между несколькими локальными центрами различных зон. Для сравнения различных способов управления информационными потоками и оценки эффективности сетей связи необходимо ввести понятие качества функционирования сети, под которым понимают совокупность ее свойств, обусловливающих способность сети удовлетворять определенные потребности в передаче информации в соответствии с назначением. Свойства сети проявляются при ее эксплуатации. Надежность, помехоустойчивость, оперативность, стабильность, экономичность, связность являются характерными свойствами сетей. Для оценки качества сетей используют показатели

качества — количественные характеристики различных свойств (коэффициент готовности, вероятность появления ошибки, время передачи информации и т. п.).

Чтобы учесть сразу ряд свойств, вводят комплексные показатели качества сетей, например время задержки сообщений в сети. Эта величина определяет скорость передачи информации, влияние помех на передачу сообщений, связность (наличие обходных путей) и надежность сети (при отказе узлов и каналов связи нужного направления время задержки сообщений растет) и многие другие свойства.

Наиболее общим является *интегральный показатель качества сети* — комплексный показатель качества сети, отражающий соотношение суммарного полезного эффекта от эксплуатации сети и суммарных затрат на ее создание и обслуживание. Выбор критерия качества сети — один из наиболее важных интуитивных элементов теории сетей. Для рассмотрения сущности задач анализа и синтеза сетей, оценки их эффективности введем следующий комплексный показатель качества:

$$\theta=\theta(\Gamma, I, C, H, \pi, V_{\text{п}}, P_{\text{пр}}, N_{\text{ин}}, \ldots,), \qquad (1.20)$$

где Г — матрица, определяющая размещение источников и потребителей информации; I — матрица интенсивностей потоков информации; C — матрица пропускных способностей ветвей; H — характеристика информационной связности сети, в общем случае учитывающая надежность ее узлов и каналов связи; π — матрица, количественно определяющая помехоустойчивость передачи информации по различным ветвям; $V_{\text{п}}$, $P_{\text{пр}}$, $N_{\text{ин}}$ — соответственно алгоритмы выбора кратчайшего пути, соблюдения приоритета и накопления информации в узлах связи.

Совокупность этих алгоритмов для всех узлов связи в целом образует общий алгоритм управления информационными потоками в сети. Указанные в выражении (1.20) характеристики являются основными и в значительной мере определяют качество работы сети. В зависимости от конкретного назначения сети набор этих характеристик может изменяться. Система управления сетью должна работать по такому алгоритму, чтобы обеспечивалось экстремальное значение показателя качества при наличии реально существующих ограничений технико-экономического характера. Например, если в роли функционала качества сети выступает среднее время задержки сообщений, то алгоритм управления сетью должен обеспечивать минимальное среднее время задержки сообщений. Построение такого алгоритма усложняется тем, что параметры I, C, H, π и другие в процессе эксплуатации изменяются случайным образом и по существу являются случайными процессами. Кроме того, установление самой связи между этими параметрами в виде соотношения (1.20) также представляет серьезную задачу.

Сущность задач анализа качества работы сети заключается в выборе показателя качества, установлении зависимости типа (1.20)

между параметрами сети и функционалом качества и анализе влияния характеристик этих параметров на функционал качества. Если анализируется система управления, то исследуется влияние структуры и параметров алгоритмов $V_п$, $P_{пр}$, $N_{ин}$ на поведение функционала качества при заданных характеристиках Г, H, I, C, π. Сущность задач оптимального синтеза сетей заключается в следующем. Известны матрицы Г, H, I, требуется выбрать функционал качества Q работы сети, установить связь этого функционала со структурой и параметрами как самой сети (C, π), так и ее системы управления ($V_п$, $P_{пр}$, $N_{ин}$). При структурной оптимизации отыскивают оптимальную структуру сети (оптимальные значения C и π) и оптимальные структуры алгоритмов $V_п$, $P_{пр}$ и $N_{ин}$, доставляющие экстремум показателю качества при наличии реально существующих ограничений технико-экономического характера. При параметрической оптимизации для выбранных структур сети и алгоритмов управления отыскивают оптимальные значения параметров этих структур.

По характеру изменения во времени структуры и параметров алгоритмов управления можно выделить сети со статическим и с динамическим управлением. Если структура и параметры алгоритмов управления постоянны, то сети называют сетями со статическим управлением. В сетях с динамическим управлением структура и параметры алгоритмов управления изменяются с изменением ситуации на сети. Если алгоритмы управления приспосабливаются к реальной ситуации, их называют адаптивными. Способы динамического управления и адаптивные алгоритмы находятся в стадии разработки и исследований.

Из постановок задач анализа и синтеза сетей следует, что их решение в полном объеме представляет сложный и трудоемкий процесс. Эти задачи являются предметом фундаментальных научных и технических исследований. Результаты исследований излагаются в курсе «Теория сетей электросвязи».

Контрольные вопросы

1. Что такое методы уплотнения линий связи?
2. Назовите особенности задач анализа и синтеза многоканальных систем.
3. В чем основная проблема теории многоканальных систем?
4. Укажите основные элементы сетей связи и их назначение.
5. Как образуется вторичная сеть связи?
6. Укажите назначение системы коммутации и системы управления.
7. Дайте определение интенсивности информационного потока, емкости канала (узла).
8. Назовите признаки классификации сетей.
9. Как определяют качество функционирования сети?
10. Какие характеристики сети определяют ее качество?
11. В чем сущность задач анализа сетей?
12. В чем сущность задач оптимального синтеза сетей?
13. В чем сущность статического и динамического управления?

1.8. ВЗАИМОСВЯЗЬ И ПРАКТИЧЕСКОЕ ИСПОЛЬЗОВАНИЕ РЕЗУЛЬТАТОВ ТЕОРИИ ИНФОРМАЦИИ И ПЕРЕДАЧИ СИГНАЛОВ

Конечная цель решения задач теории информации и передачи сигналов — создание оптимальных объектов информационной техники. С этой позиции и рассмотрим взаимосвязь шести основных классов задач (см. п. В.3) и практическое использование получаемых решений. Поставим в соответствие каждому классу задач вершину графа, отражающего взаимосвязь задач (рис. 1.4). Через γ_{ij} обозначим те результаты, которые получают из решения задач i-го класса и используют для постановки и решения задач j-го класса. Тогда ветви графа будут указывать взаимосвязи задач, а стрелки — направления передачи информации γ_{ij}.

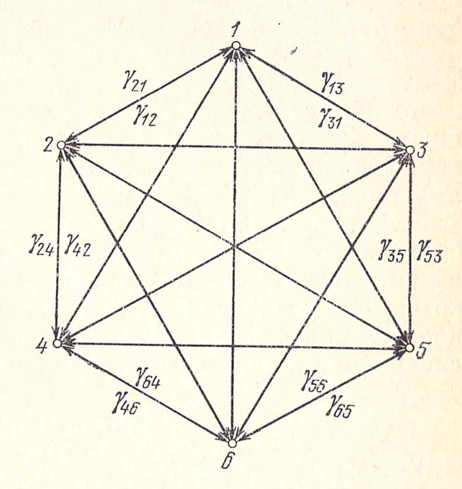


Рис. 1.4. Граф взаимосвязи основных классов задач

Анализ показывает, что полученный граф является сильно связным из-за того, что для решения задач одного класса, как правило, полезно иметь решения задач других классов. Например, при решении задач первого класса — задач анализа и синтеза сигналов — необходимо знать характеристики каналов и помех (результаты решения задач второго класса), информационные характеристики источников сообщений (результаты решения задач третьего класса), характеристики способов обработки сигналов (результаты решения задач четвертого класа), характеристики кодов (результаты решения задач пятого класа) и, конечно же, характеристики объекта информационной техники, в котором эти сигналы будут использоваться (результаты решения задач шестого класса).

Поэтому процедура решения задач оптимального синтеза объектов информационной техники носит сложный характер и требует взаимосвязанного рассмотрения постановок и результатов решения задач всех классов. По существу эта процедура является многоступенчатой и итерационной на всех стадиях проектирования и производства.

Для решения задач оптимального синтеза на первом этапе используют аналитический подход и методы параметрического анализа: задачи одного класса решают, варьируя результаты решения задач других классов, которые используют как исходные данные. На втором этапе применяют синтетический подход: когда приближенные решения задач всех необходимых классов получе-

ны, эти решения снова уточняют, но уже с позиций решения задачи оптимального синтеза в целом. Такое уточнение позволяет выяснить критичность полученных оптимальных решений к погрешностям исходных данных и областям изменения существенных переменных. Итерационную процедуру прекращают тогда, когда получают удовлетворительные решения задач шестого класса, так как именно для их правильного решения ставятся и решаются задачи первых пяти классов.

Наличие сильных взаимосвязей между результатами решения задач различных классов требует применения при проектировании объектов информационной техники системного подхода. Этот подход включает четыре взаимосвязанных этапа общей итерационной процедуры: разработка принципов действия аппаратуры и методов их инженерной реализации; анализ технико-экономических показателей качества аппаратуры, выполненный на основе различных методов; оптимизация решений относительно применяемых принципов действия, методов их технической реализации, способов конструирования и производства; анализ эффективности различных оптимальных решений.

Чтобы правильно формулировать задачи различных классов, определять необходимые исходные данные, учитывать требуемые взаимосвязи, выбирать методы решения и получать те результаты, которые позволяют успешно решать сложные задачи шестого класса, при постановке и решении задач первых пяти классов следует постоянно помнить о конечной цели.

Контрольные вопросы

1. Какая конечная цель решения задач теории информации и передачи сигналов?
2. Можно ли независимо решать задачи различных классов? Из каких соображений выбирают и учитывают необходимые взаимосвязи задач?
3. Какие подходы используют при рассмотрении задач различных классов?
4. В чем сущность системного подхода к решению задач оптимального синтеза объектов информационной техники?

1.9. ВЫВОДЫ

1. Математические модели сообщений, сигналов и помех разрабатывают для анализа их свойств и синтеза сигналов с требуемыми свойствами. При решении этой задачи используют методы теории вероятностей, теории случайных процессов, математической статистики, функционального анализа и др. Формирование сообщений рассматривают как случайный выбор реализаций случайных величин, последовательностей и процессов.

2. Для передачи информации используют сигналы, имеющие параметры селекции и информационные параметры. Параметры селекции позволяют выделить полезный сигнал из множества других сигналов и помех, а информационные параметры служат для переноса информации. Управление информационными параметрами (модуляцию) выполняют в соответствии с законом изме-

нения передаваемого сообщения. Различают непрерывные, импульсные и цифровые (дискретные) методы модуляции. Основные задачи исследования методов модуляции: анализ и синтез операторов модуляции и демодуляции, анализ свойств модулированных сигналов, синтез сигналов-переносчиков с требуемыми свойствами и др.

3. Разработка математических моделей каналов передачи информации направлена на определение структуры и параметров операторов преобразования сигналов в каналах, анализ свойств каналов и искажений сигналов в каналах, синтез каналов с требуемыми свойствами. Распространенной является модель канала в виде эквивалентного четырехполюсника, которая позволяет решать многие задачи анализа и синтеза каналов методами теории радиотехнических цепей и статистической радиотехники [6—8, 15].

4. Разработка методов расчета количества информации в сообщениях, избыточности сообщений, энтропии и производительности источника сообщений, скорости передачи информации и пропускной способности канала позволяет анализировать свойства различных источников сообщений и каналов, согласовывать источники с каналами, определять влияние помех на передачу информации, создавать коды, обладающие полезными свойствами, и т. п.

5. Из-за воздействия помех полученные сообщения отличаются от передаваемых. Поэтому разработка методов анализа и повышения помехоустойчивости передачи информации — актуальная проблема теории информации и передачи сигналов. Основными задачами этой проблемы являются выбор и обоснование критериев верности, анализ и синтез помехоустойчивых методов и алгоритмов передачи информации, создание технических средств, обеспечивающих требуемую помехоустойчивость.

6. Для повышения скорости и верности передачи информации разрабатывают методы эффективного и корректирующего кодирования. Как правило, задачи анализа и синтеза кодов, создания устройств кодирования и декодирования (кодеров, декодеров) решают тогда, когда возможности других методов повышения скорости и верности передачи информации исчерпаны.

7. Разработка методов уплотнения линий связи и методов оптимального управления информационными потоками в сетях связи направлена на повышение качества передачи информации. Задачи анализа и синтеза многоканальных систем и сетей связи являются предметом научных и технических исследований в теории многоканальной связи и теории сетей связи. Исследования проводят на основе достижений теории информации и передачи сигналов.

8. Главная цель решения задач теории информации и передачи сигналов — создание оптимальных систем и сетей связи, информационно-измерительных систем и др. Поэтому все задачи направлены на создание теоретической основы оптимального производства и эксплуатации изделий информационной техники.

Глава 2

МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ СИГНАЛОВ И ПОМЕХ

2.1. ЭЛЕМЕНТЫ ОБОБЩЕННОЙ СПЕКТРАЛЬНОЙ ТЕОРИИ СИГНАЛОВ

Обобщенная спектральная теории сигналов объединяет методы математического описания сигналов и помех. Эти методы позволяют: обеспечить требуемую избыточность сигналов, улучшить фильтрацию сигналов на фоне помех, облегчить анализ и синтез систем передачи информации (в том числе и нелинейных), улучшить синхронизацию в системах связи, создать аналоговые фильтры без индуктивностей и цифровые фильтры, повысить быстродействие цифровой обработки сигналов, решить многие другие практически важные задачи.

Обобщенной спектральной теорией сигналов называют совокупность методов представления сигналов в виде (1.3). Наибольшее распространение получили методы, использующие представления сигналов в виде колебаний (функций времени) и в виде спектрального разложения на синусоидальные и косинусоидальные составляющие (преобразования Фурье). Разложения по гармоническим функциям составляют классическую спектральную теорию. Обобщенная спектральная теория исследует общие закономерности спектрального анализа для различных систем базисных функций и рассматривает особенности выбора базисных систем при решении задач передачи и обработки сигналов.

Представление (1.3) называют разложением сигнала по системе базисных функций. К системе базисных функций предъявляют следующие основные требования: для любого сигнала ряд должен сходиться, функции $\psi_k(t)$ должны иметь простую аналитическую форму; коэффициенты a_k должны вычисляться относительно просто.

Этим трем условиям удовлетворяют системы ортогональных функций. Условие ортогональности функций имеет вид

$$\int_{t_1}^{t_2} \psi_i(t)\,\psi_k(t)\,dt = 0, \quad i \neq k. \tag{2.1}$$

При $i = k$

$$\int_{t_1}^{t_2} \psi^2{}_k(t)\,dt = c_k. \tag{2.2}$$

Число c_k называют нормой базисной функции $\psi_k(t)$.

Каждую базисную функцию можно пронормировать по ее норме, тогда нормированная базисная функция

$$\varphi_k(t)=\psi_k(t)/\sqrt{c_k}. \tag{2.3}$$

Новая система $\{\varphi_k(t)\}$ удовлетворяет не только условию ортогональности, но и условию нормировки, которое имеет вид

$$\int_{t_1}^{t_2}\varphi_i(t)\,\varphi_k(t)\,dt=\delta_{ik}, \tag{2.4}$$

где

$$\delta_{ik}=\begin{cases}0, & i\neq k\\ 1, & i=k\end{cases}$$

— символ Кронекера. Систему $\{\varphi_k(t)\}$ называют ортонормированной.

Рассмотрим, как определяются коэффициенты a_k при разложении сигнала по системе ортонормированных функций. Представим сигнал в виде

$$s(t)=\sum_{k=0}^{\infty}a_k\varphi_k(t),\quad t\in[t_1,\ t_2]. \tag{2.5}$$

Умножив обе части уравнения (2.5) на $\varphi_i(t)$ и проинтегрировав на интервале $[t_1,\ t_2]$, получим

$$\int_{t_1}^{t_2}s(t)\,\varphi_i(t)\,dt=\sum_{k=0}^{\infty}a_k\int_{t_1}^{t_2}\varphi_k(t)\,\varphi_i(t)\,dt.$$

Из условия ортонормированности (2.4) следует, что в правой части полученного уравнения все интегралы при $i\neq k$ будут равны 0, а при $i=k$ один интеграл равен 1, следовательно,

$$a_k=\int_{t_1}^{t_2}s(t)\,\varphi_k(t)\,dt. \tag{2.6}$$

Ортогональное разложение (2.5) называют обобщенным рядом Фурье, а коэффициенты, определяемые (2.6), — обобщенными коэффициентами Фурье. Ортонормированные функции удовлетворяют всем трем указанным ранее условиям. Выбор базисных ортонормированных функций — одна из ответственных задач, ее решение существенно зависит от характера преобразований сигналов в системе.

Коэффициенты a_k представляют собой эффективные значения составляющих спектра (обобщенных гармоник), поэтому средняя мощность сигнала, выделяемая на сопротивлении 1 Ом,

$$P=\frac{1}{t_2-t_1}\int_{t_1}^{t_2}s^2(t)\,dt=\sum_{k=0}^{\infty}a_k^2. \tag{2.7}$$

Соотношение (2.7) называют равенством Парсеваля. Из этого равенства следует, что мощность сигнала равна сумме мощностей всех составляющих его спектра.

Определим коэффициенты, минимизирующие погрешность ортогонального разложения. Используем среднеквадратическую погрешность

$$\varepsilon^2 = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \left[s(t) - \sum_{k=0}^{\infty} a_k \varphi_k(t) \right]^2 dt. \tag{2.8}$$

Для минимизации ε^2 необходимо решить систему уравнений $\partial\varepsilon^2 / \partial a_k = 0$, $k = 0, 1, 2, \ldots$; при условии, что $\partial\varepsilon^2 / \partial a^2_k > 0$, найти из решения $a_{k\text{опт}}$; подставить эти значения коэффициентов в соотношение (2.8) и определить

$$\varepsilon^2_{\text{мин}} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} s^2(t)\, dt - \sum_{k=0}^{\infty} a^2_{k\,\text{опт}}. \tag{2.9}$$

Эту задачу решил Фурье. Он показал, что оптимальными являются коэффициенты, определяемые по формуле (2.6). Если число членов ряда $n < \infty$, то имеется некоторая среднеквадратическая погрешность, из-за которой

$$\frac{1}{t_2 - t_1} \int_{t_1}^{t_2} s^2(t)\, dt > \sum_{k=0}^{n} a^2_{k\text{опт}}.$$

Если $n \to \infty$, то это неравенство вырождается в равенство Парсеваля (2.7), следовательно, в этом случае $\varepsilon^2_{\text{мин}} \to 0$ Таким образом, бесконечный ряд дает адекватное в среднеквадратическом смысле ортонормированное разложение сигнала.

Для реальных сигналов всегда можно указать такое, обычно небольшое n, при котором 80—90% мощности сигнала заключено в гармониках с номерами $k \leqslant n$. Поэтому ряды, используемые на практике, конечны, а число членов ряда определяет допустимые среднеквадратические погрешности. Относительную погрешность разложения определяют как отношение мощности $\Delta P(n)$ ошибки аппроксимации к мощности P самого сигнала:

$$\delta(n) = \frac{\Delta P(n)}{P} = \sum_{k=n+1}^{\infty} a^2_k \Big/ \sum_{k=0}^{\infty} a^2_k. \tag{2.10}$$

Величина $\Delta P(n)$ — та часть мощности сигнала, которая оказывается за пределами используемой полосы частот и не учитывается при восстановлении сигнала. По величине допустимой относительной погрешности δ_0 из соотношения $\delta(n) = \delta_0$ нетрудно определить число n удерживаемых членов ряда.

В качестве базисных используют системы ортогональных функций Бесселя, Хаара, Уолша, системы ортогональных полиномов Лежандра, Чебышева, Эрмита, Лаггера и др. Примеры ортогональных разложений по таким функциям и полиномам рассмотрены в [7].

Реальные сигналы всегда ограничены во времени и имеют неограниченный спектр. В теории для удобства изучения сигналы часто рассматривают не на конечном интервале $[t_1, t_2]$, а на полубесконечном $[0, \infty]$ или на бесконечном $[-\infty, \infty]$. Для определенности начало отсчета совмещают с началом сигнала или с серединой. Если сигнал имеет конечную длительность $T=t_2-t_1$, его рассматривают или на интервале $(0, T)$ или на интервале $(-T/2, T/2)$. Реальные сигналы являются случайными. Несмотря на это, в теории часто рассматривают сигналы, полностью известные в любой момент времени. Как уже отмечалось, такие сигналы называют детерминированными. Теория детерминированных сигналов как теория первого приближения удобна для решения простейших практических задач и полезна для развития теории случайных процессов.

Для изучения взаимосвязей сигналов используют такие характеристики, как взаимная энергия сигналов и взаимная мощность. *Взаимной энергией двух сигналов* называют величину

$$E_{12}=\int\limits_{-T/2}^{T/2} s_1(t)\, s_2(t)\, dt, \tag{2.11}$$

взаимной мощностью — величину

$$P_{12}=\frac{1}{T}\int\limits_{-T/2}^{T/2} s_1(t)\, s_2(t)\, dt. \tag{2.12}$$

Различают *сигналы, ортогональные по энергии,* когда $E_{12}=0$, и *ортогональные по мощности*, когда $P_{12}=0$. Для ортогональных сигналов средняя мощность и энергия суммы обладают свойством аддитивности, т. е.

$$P=\frac{1}{T}\int\limits_{-T/2}^{T/2} [s_1(t)+s_2(t)]^2\, dt=P_1+P_2,$$

$$E=\int\limits_{-T/2}^{T/2} [s_1(t)+s_2(t)]^2\, dt=E_1+E_2.$$

Сигналы, ортогональные по мощности, образуют более широкий класс, частью которого являются сигналы, ортогональные по энергии. Из ортогональности по энергии всегда следует ортогональность сигналов по мощности, но не наоборот. Только на конечном интервале, когда $T<\infty$, условия ортогональности по мощности и ортогональности по энергии выполняются одновременно.

Следовательно, ортогональность сигналов тесно связана с интервалом их определения. Напомним, что энергия сигналов измеряется в ватт-секундах (ватт на герц), а средняя мощность — в ваттах.

Взаимная энергия и взаимная мощность характеризуют степень сходства сигналов. Если два сигнала полностью совпадают, то $P_{12}=P_{21}=P$, где P — мощность сигналов. Такие сигналы называют полностью когерентными. Для ортогональных по мощности сигналов $P_{12}=P_{21}=0$, следовательно, ортогональные сигналы полностью некогерентны. Если $0<P_{12}<P$ или $0<P_{21}<P$, то сигналы называют частично-когерентными.

Контрольные вопросы

1. В чем сущность ортогонального разложения сигнала?
2. Запишите условия ортогональности и ортонормированности функций.
3. Как определяют коэффициенты ортонормированных разложений?
4. Что такое среднеквадратическая погрешность ортонормированного разложения?
5. Как определяют относительную погрешность ортонормированного разложения?
6. Как определяют ортогональность и когерентность сигналов?

2.2. ОРТОГОНАЛЬНЫЕ РАЗЛОЖЕНИЯ КОТЕЛЬНИКОВА

Ортогональные разложения Котельникова для непрерывных сигналов с ограниченными и полосовыми спектрами, так же как и преобразования Фурье для периодических и непериодических сигналов, являются примерами наиболее распространенного применения ортогональных разложений. Рассмотрим основные особенности ортогональных разложений Котельникова.

2.2.1. Ортогональное разложение Котельникова для непрерывных сигналов с ограниченными спектрами имеет важное значение, так как позволяет представлять непрерывные сигналы в виде импульсных последовательностей. Теоретической основой разложения служит *теорема Котельникова (теорема отсчетов):* любая непрерывная функция $s(t)$, не содержащая частот выше F, полностью определяется последовательностью своих значений в моменты времени, отстоящие друг от друга на интервал $\Delta t=1/2F$. С доказательством теоремы можно познакомиться в учебниках [1—3]. Общее число отсчетов для сигнала длительностью T $n=T/\Delta t=2FT=\nu$, т. е. равно базе сигнала.

Ортогональное разложение Котельникова для сигнала $s(t)$, спектр которого лежит в интервале $[0, F]$, имеет вид [1—3]:

$$s(t)=\sum_{k=-\infty}^{\infty} s(k\Delta t)\,\frac{\sin 2\pi F\,(t-k\Delta t)}{2\pi F\,(t-k\Delta t)}, \qquad (2.13)$$

где $s(k\Delta t)=s_k$ — отсчет сигнала в момент времени t_k; $\left\{\frac{\sin 2\pi F\,(t-k\Delta t)}{2\pi F\,(t-k\Delta t)}\right\}$ — базисная система ортогональных функций с общей нормой $1/2F$; $\Delta t=1/2F$ — интервал дискретизации, равный норме базисных функций. Функции $g_k(t)=$ $=\sin 2\pi F(t-k\Delta t)/2\pi F(t-k\Delta t)$ — называют функциями отсчетов, а значения s_k — отсчетами. График функции отсчетов приведен на рис. 2.1.

Ортогональность функций отсчетов легко проверяется непосредственно путем вычисления интеграла

$$\int_{-\infty}^{\infty} \frac{\sin 2\pi F(t-k\Delta t)}{2\pi F(t-k\Delta t)} \frac{\sin 2\pi F(t-j\Delta t)}{2\pi F(t-j\Delta t)} dt = \begin{cases} 1/2F, & j=k, \\ 0, & j \neq k. \end{cases}$$

Интервал дискретизации, как видим, не превышает половины периода наиболее высокой частоты спектра сигнала.

Из равенства Парсеваля (2.7) следует, что энергия непрерывного сигнала с ограниченным спектром определяется через отсчеты:

$$E = \int_{-\infty}^{\infty} s^2(t)\, dt = \frac{1}{2F} \sum_{-\infty}^{\infty} s^2{}_k. \tag{2.14}$$

Еще раз следует подчеркнуть, что в природе нет сигналов, которые имеют одновременно ограниченную длительность и ограниченный спектр. Однако в инженерных расчетах необходимо учитывать ту часть спектра, в которой сосредоточено 80—95% энергии сигнала. Поэтому чаще всего большинство сигналов рассматривают как сигналы с ограниченными спектрами. Если сигнал существует на интервале $[t_1, t_2]$, то число членов ряда (число отсчетов) выбирают равным $(t_2-t_1)/\Delta t = 2TF = \nu$. Например, у телефонного сигнала 95% энергии сосредоточено в полосе частот 300—3400 Гц. Если верхней частотой считать $F=$ $=3400$ Гц, то частота дискретизации этого сигнала $2F=6800$ Гц.

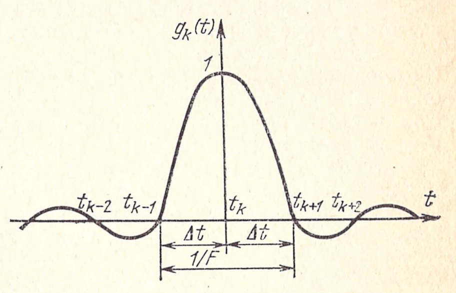


Рис. 2.1. График функции отсчетов

Достоинства ортогонального разложения Котельникова (2.13) следующие: базисная система ортогональных функций выбрана так, что ряд (2.13) носит формальный характер, т. е. в любой момент времени отсчета t_k этот ряд дает только одно значение s_k, все остальные составляющие ряда вырождаются в нуль; коэффициенты ряда (2.13) нет необходимости вычислять: они определяются непосредственно путем измерения значений сигнала или из его аналитической формы; зная длительность сигнала T и граничную частоту F, легко определить требуемое число отсчетов $n=2TF$ и энергию сигнала из (2.14); относительная простота аппаратурной реализации как разложения (дискретизации) непрерывного сигнала в импульсную последовательность, так и последующего его восстановления.

На последней особенности, имеющей важное практическое значение, целесообразно остановиться более подробно. Для этого рассмотрим физический смысл разложения Котельникова. Каждый член суммы разложения (2.13) представляет собой отклик идеального фильтра нижних частот g_k (рис. 2.1) с частотой среза F на очень короткий импульс, приходящий в момент времени $t_k=k\Delta t$ и имеющий площадь, равную $s(k\Delta t)$. Поэтому при дискретной передаче сигнала $s(t)$ с ограниченным спектром по каналу связи необходимо через равные интервалы времени Δt брать отсчеты мгновенных значений сигнала и передавать по каналу последовательность достаточно коротких импульсов длительностью $\tau(\tau/\Delta t \ll 1)$, амплитуда которых A_k в момент времени $t_k=k\Delta t$ выбирается так, чтобы $A_k\tau=s(k\Delta t)$.

В приемном устройстве выделенная последовательность видеоимпульсов пропускается через фильтр нижних частот, на выходе которого восстанавливается переданный непрерывный сигнал. Длительность импульсов τ может быть сколь угодно малой, но реально выбирается исходя из полосы прозрачности канала связи. Частота дискретизации (тактовая частота), как уже отмечалось, равна $2F$.

2.2.2. Ортогональное разложение Котельникова для непрерывных сигналов с полосовыми спектрами. Если сигнал $s(t)$ имеет полосовой спектр и ширина полосы спектра $\Delta F_1 = f_2 - f_1$, то такой непрерывный сигнал можно представить в виде следующего ортогонального разложения [1—3]

$$s(t) = \sum_{k=-\infty}^{\infty} s\left(\frac{k}{\Delta F_1}\right) \frac{\sin \pi \Delta F_1 (t - k/\Delta F_1)}{\pi \Delta F_1 (t - k/\Delta F_1)} \times$$
$$\times \cos\left[\omega_0 \left(t - \frac{k}{\Delta F_1}\right) + \varphi\left(\frac{k}{\Delta F_1}\right)\right]. \qquad (2.15)$$

Здесь среднее значение угловой частоты спектра сигнала $\omega_0 = 2\pi(f_1+f_2)/2$; $\Delta t = 1/2\Delta F_1$; $s(k/\Delta F_1)$, $\varphi(k/\Delta F_1)$ — отсчеты амплитуды и фазы сигнала в моменты времени $t_k = k\Delta t$. Следовательно, для сигналов с полосовыми спектрами необходимо через интервал дискретизации отсчитывать не только мгновенные значения амплитуд, но и мгновенные значения фаз. Так, например, выполняют дискретизацию однополосных сигналов, типичных сигналов с полосовыми спектрами. Если телефонный сигнал рассматривают как сигнал с полосовым спектром, частоту дискретизации выбирают, равной 6200 Гц (см. пример в п. 2.2.1).

Основные особенности ортогонального разложения Котельникова (2.15) следующие: базисная система включает совокупность ортогональных функций отсчетов, каждая из которых представляет модулированное колебание с несущей частотой ω_0 и огибающей, определяемой функцией типа $g_k(t)$; помимо отсчетов амплитуд, берутся и отсчеты фаз; если длительность сигнала T, то количество отсчетных точек $n = T/\Delta t = 2T\Delta F_1$. Ортогональные разложения Котельникова являются теоретической основой большинства методов дискретной передачи непрерывных сигналов. Они позволяют с единых позиций рассматривать передачу как дискретных, так и непрерывных сигналов.

Контрольные вопросы

1. Сформулируйте теорему Котельникова.
2. Запишите и поясните сущность ортогонального разложения Котельникова для непрерывных сигналов с ограниченными спектрами.
3. Укажите достоинства разложения (2.13).
4. Поясните физический смысл разложения (2.13).
5. Запишите и поясните сущность ортогонального разложения Котельникова для непрерывных сигналов с полосовыми спектрами.
6. В чем практическая ценность разложений Котельникова?

2.3. КОРРЕЛЯЦИОННЫЕ И СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ СИГНАЛОВ И ПОМЕХ

Корреляционные и спектральные характеристики случайных процессов детально рассматриваются в статистической радиотехнике. Поэтому здесь целесообразно кратко систематизировать лишь те сведения о характеристиках, которые необходимы для изложения последующих разделов.

Корреляционная функция $K(t_1, t_2)$ показывает взаимосвязь (корреляцию) значений $X(t_1) = X_1$ и $X(t_2) = X_2$ случайного процесса $X(t)$ в моменты времени t_1 и t_2 [7, 15]:

$$K(t_1, t_2) = \iint_{-\infty}^{\infty} (x_1 - m_1)(x_2 - m_2) f_2(x_1, x_2; t_1, t_2)\, dx_1 dx_2, \qquad (2.16)$$

где $m_i = \int_{-\infty}^{\infty} x_i f_1(x_i, t_i)\, dx_i$ — математическое ожидание процесса в моменты t_i;

$f_1(x_i, t_i)$, $f_2(x_1, x_2; t_1, t_2)$ — одномерная и двумерная плотности распределения $X(t)$. Если $K(t_1, t_2)=0$, сечения X_1 и X_2 — некоррелированны.

Для стационарных случайных процессов $m_1=m_2=m$, а корреляционная функция зависит только от $\tau=t_2-t_1$ [1—3], т. е.

$$K(\tau) = \iint\limits_{-\infty}^{\infty} (x_1 - m)(x_2 - m) f_2(x_1, x_2; \tau)\, dx_1 dx_2. \qquad (2.17)$$

Часто используют нормированную корреляционную функцию $\rho(\tau)= =K(\tau)/K(0)$, где $K(0)$ является дисперсией $\mathscr{D}$ процесса. Функция $\rho(\tau)$ обладает следующими свойствами: $\rho(0)=1$; $\rho(0)\geqslant|\rho(\tau)|$; $\rho(\tau)=\rho(-\tau)$; если $m=0$, $\lim\limits_{\tau\to\infty} \rho(\tau) = 0$.

Интегральной характеристикой времени корреляции сечений процесса служит интервал корреляции

$$\Delta\tau = 2 \int\limits_0^{\infty} |\rho(\tau)|\, d\tau. \qquad (2.18)$$

Если сечения отстоят друг от друга на расстояние большее $\Delta\tau$, для инженерных расчетов их считают некоррелированными.

Операцию определения корреляционных функций с помощью интегралов (2.16), (2.17) называют усреднением по множеству (по ансамблю). Для сокращения записи обозначим ее через $M[\cdot]$. Например, (2.17) сокращенно удобно записывать так:

$$K(\tau) = M[(X_1-m)(X_2-m)].$$

В экспериментальных исследованиях характеристики случайных процессов получают чаще всего усреднением по времени. Эту операцию обозначим через $\langle \cdot \rangle$. Оценка математического ожидания процесса по j-й реализации длительностью T

$$m^*_j(T) = \langle X_j(t) \rangle = \frac{1}{T} \int\limits_{-T/2}^{T/2} x_j(t)\, dt. \qquad (2.19)$$

Оценка корреляционной функции

$$K^*_j(\tau, T) = \langle [X_j(t) - m^*][X_j(t+\tau) - m^*] \rangle = \frac{1}{T} \int\limits_{-T/2}^{T/2} [x_j(t) - m^*][x_j(t + \tau) - m^*]\, dt. \qquad (2.20)$$

(Звездочка указывает, что оценки являются случайными величинами, которые зависят от номера j выбранной реализации и длительности интервала наблюдения T.)

Стационарные случайные процессы, у которых средние по времени совпадают в предельном смысле со средними по множеству, называют *эргодическими*, а такое свойство процессов — *эргодичностью*. Например, для эргодических процессов для любого j с вероятностью единица выполняются условия

$$M[X(t)] = \lim_{T\to\infty} \langle X_j(t) \rangle, \quad j = 1, 2, \ldots,$$
$$M[(X_1 - m)(X_2 - m)] = \lim_{T\to\infty} \langle [X_j(t) - m^*][X_j(t+\tau) - m^*] \rangle. \qquad (2.21)$$

Эргодичность процессов имеет важное практическое значение потому, что наблюдение за большим числом реализаций случайного процесса можно заменить наблюдением всего лишь за одной, но достаточно продолжительной реализацией. Полученные таким образом характеристики процесса (математическое ожидание,

дисперсия, корреляционная функция, спектральная плотность и др.) будут с достаточной для инженерных расчетов точностью совпадать с теми, которые получают путем обработки большого числа реализаций.

Корреляционные и спектральные характеристики случайного процесса связаны соотношениями Хинчина — Винера [1—3, 15]:

$$\Omega(\omega) = \int_{-\infty}^{\infty} K(\tau)\, e^{-i\omega\tau} d\tau = 2\int_{0}^{\infty} K(\tau) \cos \omega\tau d\tau,$$

$$K(\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \Omega(\omega)\, e^{i\omega\tau}\, d\omega = \frac{1}{\pi}\int_{0}^{\infty} \Omega(\omega) \cos \omega\tau d\omega, \tag{2.22}$$

где $\Omega(\omega)$ — спектральная плотность случайного процесса. Эти соотношения являются преобразованиями Фурье для случайных процессов. Особенность их в том, что в интегралах фигурируют не сами процессы, а их корреляционные функции.

Из (2.22) следует, что дисперсия процесса

$$\mathscr{D} = K(0) = \frac{1}{\pi}\int_{0}^{\infty} \Omega(\omega)\, d\omega. \tag{2.23}$$

Интегральной характеристикой спектральной плотности процесса служит ширина спектра

$$\Delta\omega = \frac{1}{\Omega_m}\int_{0}^{\infty} \Omega(\omega)\, d\omega = \pi\mathscr{D}/\Omega_m. \tag{2.24}$$

Ширина спектра — это основание прямоугольника с высотой Ω_m, площадь которого равна площади под кривой $\Omega(\omega)$ (Ω_m — максимальное значение $\Omega(\omega)$).

Наряду с корреляционной функцией и спектральной плотностью используют взаимно-корреляционные функции и взаимные спектральные плотности процессов. Взаимно-корреляционная функция двух процессов $X_1(t)$ и $X_2(t)$

$$K_{12}(\tau) = M[[X_1(t) - m_1][X_2(t+\tau) - m_2]], \tag{2.25}$$

где m_1 и m_2 — математические ожидания этих процессов. Взаимная спектральная плотность

$$\Omega_{1,2}(\omega) = \int_{-\infty}^{\infty} K_{1,2}(\tau)\, e^{-i\omega\tau} d\tau = 2\int_{0}^{\infty} K_{1,2}(\tau) \cos \omega\tau d\tau. \tag{2.26}$$

Если взаимные характеристики двух процессов, полученные путем усреднения по множеству, совпадают с характеристиками, полученными путем усреднения по времени, то такие процессы называют совместно-эргодическими.

Контрольные вопросы

1. Дайте определение корреляционной функции случайного процесса.
2. Как определяют интервал корреляции?
3. Как экспериментально оценивают характеристики случайных процессов?
4. Какие случайные процессы называют эргодическими?
5. Какими соотношениями связаны корреляционные и спектральные характеристики процесса?
6. Как определить дисперсию процесса по его спектральной плотности?
7. Как определяют ширину спектра случайного процесса?
8. Дайте определение взаимно-корреляционной функции и взаимной спектральной плотности.

Наиболее распространенными моделями случайных сигналов и помех являются телеграфный сигнал, белый шум, гауссовский случайный процесс, гауссовский белый шум. Рассмотрим вероятностные характеристики этих процессов.

2.4.1. Телеграфным сигналом называют случайную видеопоследовательность прямоугольных положительных и отрицательных импульсов со случайными длительностями τ_1, τ_2 и детерминированными амплитудами σ, $-\sigma$.

Если длительности импульсов распределены по показательным законам $f_1(\tau_1)$, $f_2(\tau_2)$ с параметрами λ_1 и λ_2, т. е. $f_1(\tau_1)=\lambda_1 e^{-\lambda_1\tau_1}$, $f_2(\tau_2)=\lambda_2 e^{-\lambda_2\tau_2}$, то телеграфный сигнал является стационарным случайным процессом, который имеет показательную корреляционную функцию [1, 15]

$$K(\tau)=\sigma^2 e^{-(\lambda_1+\lambda_2)\,|\tau|}=\sigma^2 e^{-\alpha\,|\tau|}, \qquad (2.27)$$

где σ^2 — дисперсия процесса; $\alpha=\lambda_1+\lambda_2$ — параметр, значения которого полностью определяют корреляционные и спектральные свойства телеграфного сигнала.

Телеграфный сигнал обладает важным свойством. Изменением α можно в широком диапазоне изменять корреляционные и спектральные характеристики процесса. При $\alpha\to 0$ характеристики телеграфного сигнала приближаются к характеристикам постоянной составляющей, при $\alpha\to\infty$ — к характеристикам белого шума.

Определим интервал корреляции, спектральную плотность и ширину спектра телеграфного сигнала. Интервал корреляции найдем по формуле (2.18)

$$\Delta\tau=2\int_0^\infty \frac{K(\tau)}{\sigma^2}\,d\tau=\frac{2}{\alpha}. \qquad (2.28)$$

Из (2.28) следует, что чем больше α, тем меньше время корреляции процесса. При $\alpha\to 0$ $\Delta\tau\to\infty$ и процесс вырождается в детерминированный. При $\alpha\to\infty$ $\Delta\tau\to 0$ и процесс вырождается в белый шум, у которого все сечения, в том числе и соседние, не коррелированы.

Спектральную плотность телеграфного сигнала определим с помощью (2.22):

$$\Omega(\omega)=2\int_0^\infty \sigma^2 e^{-\alpha|\tau|}\cos\omega\tau\,d\tau=\frac{2\sigma^2\alpha}{\alpha^2+\omega^2}, \qquad (2.29)$$

На рис. 2.2 и 2.3 показаны графики функций $K(\tau)$ и $\Omega(\omega)$ при $\alpha_1<\alpha_2$.

В соответствии с (2.24) ширина спектра телеграфного сигнала

$$\Delta\omega=\frac{\alpha}{2\sigma^2}\int\limits_0^\infty\frac{2\sigma^2\alpha}{\alpha^2+\omega^2}d\omega=\alpha\pi,\quad \Delta F_1=\frac{\Delta\omega}{2\pi}=\frac{\alpha}{2}. \tag{2.30}$$

При $\alpha\to0$ $\Delta\omega\to0$ и процесс вырождается в постоянную составляющую. При $\alpha\to\infty$ $\Delta\omega\to\infty$ и процесс вырождается в белый шум, у которого спектральная плотность постоянная в широком диапазоне частот.

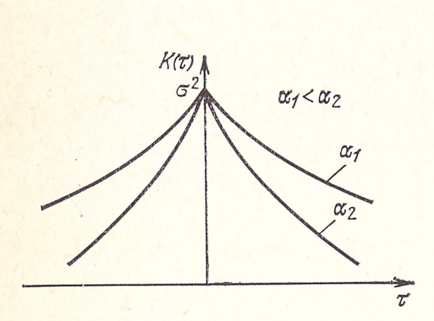


Рис. 2.2. Графики $K(\tau)$ при $\alpha_1<\alpha_2$

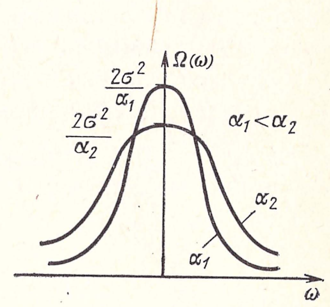


Рис. 2.3. Графики $\Omega(\omega)$ при $\alpha_1<\alpha_2$

Интересно отметить, что для телеграфного сигнала произведение

$$\Delta\tau\Delta F_1=1, \tag{2.31}$$

Отсюда следует важный вывод, что спектр случайного процесса тем шире, чем меньше интервал корреляции процесса.

2.4.2. Белый шум используют как модель наиболее тяжелого вида помехи в каналах связи. Он является стационарным случайным процессом с постоянной спектральной плотностью $\Omega(\omega)=\Omega_0$. Название «белый шум» возникло по аналогии с применяемым в оптике понятием белый свет, который содержит все цвета спектра и все спектральные составляющие которого имеют примерно одинаковую энергию.

Определяя белый шум как предельное состояние телеграфного сигнала при $\alpha\to\infty$, найдем свойства белого шума. Умножим и разделим выражение для спектральной плотности телеграфного сигнала на α. Введем спектральную плотность белого шума

$$\Omega_0=2\sigma^2/\alpha, \tag{2.32}$$

тогда

$$\Omega(\omega)=\Omega_0\alpha^2/(\alpha^2+\omega^2)=\Omega_0/(1+\omega^2/\alpha^2). \tag{2.33}$$

Из (2.33) следует, что $\lim\Omega(\omega)=\Omega_0$, если и $\sigma^2\to\infty$ так, что $\lim\limits_{\sigma^2,\,\alpha\to\infty}(\sigma^2/\alpha)=\mathrm{const}$.

Рассмотрим, как изменяются дисперсия и корреляционная функция телеграфного сигнала при $\alpha\to\infty$. Используя (2.32), выразим через α дисперсию

$$\sigma^2=0{,}5\Omega_0\alpha=\Omega_0\Delta F_1. \tag{2.34}$$

По физическому смыслу спектральная плотность — это мощность процесса, которая приходится на 1 Гц полосы частот, так как $\Omega_0=\sigma^2/\Delta F_1$. Из (2.34) следует, что $\lim\limits_{\alpha\to\infty}\sigma^2\to\infty$, т. е. мощность белого шума не ограничена. Подставив значение мощности из (2.34) в (2.27), получим

$$K(\tau)=0{,}5\Omega_0\alpha e^{-\alpha|\tau|}. \qquad (2.35)$$

Так как $\lim\limits_{\alpha\to\infty}\alpha e^{-\alpha|\tau|}\to 0$ при $|\tau|\neq 0$, то $\lim\limits_{\alpha\to\infty,\ |\tau|\neq 0}K(\tau)\to 0$. При $|\tau|\to 0$ и $\alpha\to\infty$, но так, что $\alpha|\tau|=\mathrm{const}$, $\lim\limits_{\alpha\to\infty,\ |\tau|\to 0}K(\tau)\to\infty$. Поэтому корреляционную функцию белого шума в окрестности точки $\tau=0$ аппроксимируют дельта-функцией [1, 15] и записывают ее в виде

$$K(\tau)=0{,}5\Omega_0\delta(\tau). \qquad (2.36)$$

Определим спектральную плотность белого шума через $K(\tau)$:

$$\Omega(\omega)=2\int\limits_0^\infty\frac{\Omega_0}{2}\delta(\tau)\cos\omega\tau\,d\tau=\Omega_0. \qquad (2.37)$$

Таким образом, белый шум обладает следующими свойствами: спектральная плотность белого шума постоянна, значения белого шума при любых $|\tau|\neq 0$ некоррелированы, дисперсия белого шума бесконечна.

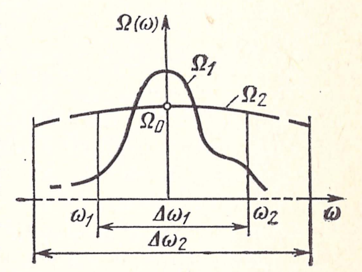


Рис. 2.4. Графики спектральных характеристик сигнала (*1*) и помехи (*2*)

Многие помехи в технике связи, вычислительной технике и других областях рассматривают как белый шум. К таким помехам относят флуктуационные шумы, помехи в многоканальных системах и сетях связи и др. Важно отметить, что белый шум является идеализацией. В природе не существует источников сигналов и помех, которые могли бы обеспечить бесконечную мощность сигналов и помех, а также генерировать реализации процессов с некоррелированными близкими отсчетами. Тем не менее, этой идеализацией можно пользоваться, если действие помехи с шириной спектра $\Delta\omega_2$ рассматривается в полосе частот $\Delta\omega_1$ полезного сигнала или системы и соблюдается условие

$$\Delta\omega_1/\Delta\omega_2\ll 1, \qquad (2.38)$$

а спектральная плотность помехи $\Omega_2(\omega)$ слабо изменяется на интервале $\Delta\omega_1$. На рис. 2.4 эти условия иллюстрируют графики спектральных характеристик сигнала (системы) и помехи. Характеристиками помехи как белого шума служат спектральная плотность Ω_0 и средняя мощность в полосе частот сигнала

$$\sigma^2=\Omega_0\Delta F_1=\Omega_0\Delta\omega_1/2\pi. \qquad (2.39)$$

Частотные составляющие помехи, которые лежат за пределами полосы пропускания системы, в инженерных расчетах можно не учитывать.

2.4.3. Гауссовский процесс. Случайный процесс, n-мерная плотность распределения которого имеет вид

$$f_n(\overline{x_1, x_n}; \overline{t_1, t_n}) = \frac{1}{\sqrt{(2\pi\sigma^2)^n A}} \exp\left(-\frac{1}{2\sigma^2 A} \sum_{i=1}^{n} \sum_{k=1}^{n} A_{ik} x_i x_k\right), \quad (2.40)$$

называют гауссовским. Здесь

$$A = \begin{vmatrix} R_{11} & R_{12} & \ldots & R_{1n} \\ R_{21} & R_{22} & \ldots & R_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ R_{n1} & R_{n2} & \ldots & R_{nn} \end{vmatrix}$$

— определитель; σ^2 — дисперсия процесса; $m=0$; $R_{ik}=K(t_i, t_k)$; A_{ik} — алгебраическое дополнение R_{ik} в A.

Для стационарного процесса $R_{ik}=R_{ki}=K(\tau)$, где $\tau=t_k-t_i$. Поэтому для гауссовского процесса по корреляционной функции можно определить плотность распределения любого порядка. Первые две плотности распределения этого процесса имеют вид:

$$f_1(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/2\sigma^2}, \quad x \in (-\infty, \infty), \quad (2.41)$$

$$f_2(x_1, x_2, \tau) = \frac{1}{2\pi\sigma^2 \sqrt{1-\rho^2(\tau)}} \exp\left[-\frac{x_1^2 + x_2^2 + 2\rho(\tau) x_1 x_2}{2\sigma^2[1-\rho^2(\tau)]}\right], \quad (2.42)$$

где нормированная корреляционная функция $\rho(\tau)=K(\tau)/\sigma^2$.

2.4.4. Гауссовский белый шум. Если гауссовский процесс является белым шумом, все n сечений его некоррелированы, $A_{ik}=1$, $A=1$, $R_{ik}=R_{ki}=\delta_{ik}$ (δ_{ik} — символ Кронекера). Поэтому плотность распределения n-го порядка для гауссовского белого шума определяют как произведение из n одномерных плотностей распределения

$$f_n(\overline{x_1, x_n}; \overline{t_1, t_n}) = \prod_{i=1}^{n} f(x_i, t_i) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} x_i^2\right). \quad (2.43)$$

Распределенное по закону Гаусса колебание образуется в результате сложения большого числа независимых или слабо коррелированных случайных колебаний.

Контрольные вопросы

1. Какие модели случайных сигналов и помех наиболее распространены?
2. Опишите свойства и характеристики телеграфного сигнала.
3. Опишите свойства и характеристики белого шума.
4. Опишите свойства и характеристики гауссовского процесса.
5. Как определяют многомерную плотность распределения гауссовского белого шума?

2.5. КАНОНИЧЕСКИЕ И НЕКАНОНИЧЕСКИЕ РАЗЛОЖЕНИЯ СЛУЧАЙНЫХ СИГНАЛОВ И ПОМЕХ

Каноническим разложением случайного процесса $X(t)$ называют ортогональное разложение типа

$$X(t)=\sum_{k=0}^{n} X_k\varphi_k(t), \quad t\in[0, T], \tag{2.44}$$

где $\{\varphi_k(t)\}$ — система ортогональных или ортонормированных функций; T — интервал наблюдения или длительность процесса; $n=2TF$, F — верхняя частота спектра процесса (если $F\to\infty$, то и $n\to\infty$); $\{X_k\}$ — система некоррелированных случайных величин.

Для стационарных случайных процессов наиболее удобны разложения по гармоническим функциям

$$X(t)=m+\sum_{k=1}^{n}(a_k\cos\omega_k t+b_k\sin\omega_k t), \tag{2.45}$$

где $m=M[X(t)]$; $M[a_k]=M[b_k]=0$ $M[a^2{}_k]=M[b^2{}_k]=\sigma^2{}_k=\mathscr{D}_k$ — дисперсия k-й гармоники; $\omega_k=k\omega_1$; $\omega_1=\pi/T$.

Корреляционная функция разложения (2.45)

$$K(\tau)=\sum_{k=1}^{n}\mathscr{D}_k\cos\omega_k\tau, \tag{2.46}$$

где $\tau=t_2-t_1$. Так как дисперсия $\mathscr{D}$ стационарного случайного процесса равна его корреляционной функции в начале координат, то

$$\mathscr{D}=K(0)=\sum_{k=1}^{n}\mathscr{D}_k. \tag{2.47}$$

Из (2.47) следует, что так же, как и для детерминированного сигнала (см. (2.7)), мощность случайного сигнала равна сумме мощностей гармонических составляющих разложения (2.45). Формула (2.47) показывает, как мощность процесса распределена по гармоникам.

Покажем, как определяют дисперсию $\mathscr{D}_k$ по корреляционной функции $K(\tau)$. Умножив почленно равенство (2.46) на $\cos\omega_j\tau$ и проинтегрировав результат по τ в пределах от 0 до T, получим

$$\int_0^T K(\tau)\cos\omega_j\tau\,d\tau=\sum_{k=1}^{n}\mathscr{D}_k\int_0^T\cos\omega_k\tau\cos\omega_j\tau\,d\tau.$$

При $k\neq j$ интеграл в правой части равен нулю, а при $k=j$ он равен $T/2$ — норме базисных функций. Поэтому

$$\mathscr{D}_k=\frac{2}{T}\int_0^T K(\tau)\cos\omega_k\tau\,d\tau, \quad k=\overline{1, n}. \tag{2.48}$$

С увеличением длительности T процесса число отсчетов n также растет, спектр частот процесса становится непрерывным. Случайные величины в (2.45) необходимо заменить бесконечно малыми случайными величинами $a(\omega)d\omega$ и $b(\omega)d\omega$, а сумму заменить интегралом по ω. Тогда

$$X(t) = m + \int_0^\infty [a(\omega)\cos\omega t + b(\omega)\sin\omega t]\, d\omega.$$

Обозначив бесконечно малую дисперсию $\mathcal{D}_k$ случайных величин $a(\omega)d\omega$ и $b(\omega)d\omega$ через $\Omega(\omega)d\omega$, после необходимых преобразований получим уже известное соотношение Хинчина — Винера (2.22) для корреляционной функции процесса. Путем предельного перехода при $T\to\infty$ аналогично получают соотношение Хинчина — Винера (2.22) для спектральной плотности.

Нестационарные случайные процессы также можно представлять в виде канонических разложений по гармоническим функциям. Однако коэффициенты таких разложений получаются уже коррелированными между собой, что неудобно для решения прикладных задач. Поэтому для ортогональных разложений нестационарных случайных процессов отыскивают другие базисные системы, которые приводят к некоррелированным коэффициентам. Примером такого разложения служит разложение Лоэва — Карунена.

Рассмотрим физический смысл интервала дискретизации в каноническом разложении Котельникова для случайных процессов. Для этого найдем связь интервала дискретизации и интервала корреляции. Ширина спектра непрерывного случайного процесса с ограниченным спектром $\Delta\omega = 2\pi F$. Так как $\Delta t = 1/2F$, то

$$\Delta t = \pi/\Delta\omega. \qquad (2.49)$$

Ширину спектра процесса выразим через его дисперсию $\mathcal{D}$ и максимальное значение спектральной плотности Ω_m из формулы (2.24), тогда

$$\Delta t = \Omega_m/\mathcal{D}. \qquad (2.50)$$

Дисперсию процесса определим через интервал корреляции и значение спектральной плотности в начале координат $\Omega(0)$ из соотношения

$$\Delta\tau = \frac{2}{\mathcal{D}} \int_0^\infty \rho(\tau)\cos\omega\tau\, d\tau \Big|_{\omega=0} = \frac{\Omega(0)}{\mathcal{D}},$$

тогда

$$\mathcal{D} = \Omega(0)/\Delta\tau. \qquad (2.51)$$

Подставив значение дисперсии из (2.51) в (2.50), найдем

$$\Delta t = \frac{\Omega_m}{\Omega(0)} \Delta\tau. \qquad (2.52$$

Соотношение (2.52) играет важную роль, оно устанавливает связь интервала дискретизации случайного процесса с интервалом корреляции и значениями спектральной плотности Ω_m, $\Omega(0)$. Так как

$$\Omega_m/\Omega(0) \geqslant 1, \tag{2.53}$$

то и отношение

$$\Delta t/\Delta\tau \geqslant 1. \tag{2.54}$$

Следовательно, если интервал дискретизации случайного сигнала выбирать в соответствии с теоремой Котельникова, то отсчеты будут некоррелированными.

Существенными недостатками канонических разложений случайных процессов являются большое число случайных переменных и низкая эффективность при анализе нелинейных систем. Поэтому разрабатывают и неканонические представления случайных процессов, например в виде нелинейной функции нескольких случайных аргументов.

Удобное неканоническое представление предложил В. И. Чернецкий. Это представление позволяет абсолютно точно в рамках корреляционной теории описывать стационарные случайные процессы в виде детерминированной функции всего лишь трех случайных аргументов. Рассмотрим основные особенности этого представления.

Если в разложении (2.45) допустить, что частота также является непрерывной случайной величиной, то любой случайный процесс $X(t)$, как доказано в теореме Чернецкого, можно представить в виде

$$X(t) = m(t) + a\cos\omega t + b\sin\omega t, \tag{2.55}$$

где $m(t) = M[X(t)]$, a, b, ω — независимые случайные величины, которые определяют по характеристикам процесса. В (2.55) сокращение числа случайных переменных и числа гармоник произведено в результате перехода к нелинейному представлению.

Если в представлении (2.55) $M[a] = M[b] = 0$, $M[a^2] = M[b^2] = \mathscr{D}$, закон распределения ω

$$f(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \rho(\tau)\, e^{-i\omega\tau} d\tau, \tag{2.56}$$

процесс $\Delta X(t) = X(t) - m(t)$ является стационарным, то с помощью (2.55) удобно описывать случайные процессы, приводимые к стационарным путем центрирования — вычитания из процесса его математического ожидания. Спектральная плотность процесса

$$\Omega(\omega) = 2\pi\mathscr{D}f(\omega), \tag{2.57}$$

следовательно, плотность распределения частоты можно найти и по спектральной плотности, если известна дисперсия процесса.

Если обозначить $a = A\cos\Phi$, $b = A\sin\Phi$ и применить формулу косинуса разности двух углов, то стационарный процесс $\Delta X(t)$

можно представить в виде гармонического колебания, которое имеет случайные амплитуду, частоту и фазу,

$$\Delta X(t) = A\cos(\omega t - \Phi), \tag{2.58}$$

где $A = \sqrt{a^2 + b^2}$; $\Phi = \operatorname{arctg}(b/a)$. Представление (2.58) для стационарных случайных процессов является удобным для решения многих инженерных задач. Например, используя (2.58), нетрудно обобщить результаты анализа всех видов модуляции детерминированной гармонической несущей на случайные стационарные несущие.

Контрольные вопросы

1. Что называют каноническим разложением случайного процесса, какое разложение наиболее удобно для стационарного процесса?
2. Как определяют корреляционную функцию разложения?
3. Как определяют дисперсии гармоник по корреляционной функции процесса?
4. Как получают из канонического разложения соотношения Хинчина — Винера?
5. Как связаны интервал дискретизации и интервал корреляции процесса?
6. Назовите особенности неканонического представления Чернецкого?

2.6. УЗКОПОЛОСНЫЕ И АНАЛИТИЧЕСКИЕ СИГНАЛЫ

Узкополосные и аналитические сигналы широко используют как модели реальных сигналов и помех. Процесс называют *узкополосным*, если

$$\Delta\omega / \omega_0 \ll 1, \tag{2.59}$$

где ширина спектра процесса $\Delta\omega = \omega_2 - \omega_1$, а средняя частота $\omega_0 = (\omega_2 + \omega_1)/2$. Реализации узкополосных процессов можно наблюдать на выходе схем, работающих на высоких и промежуточных

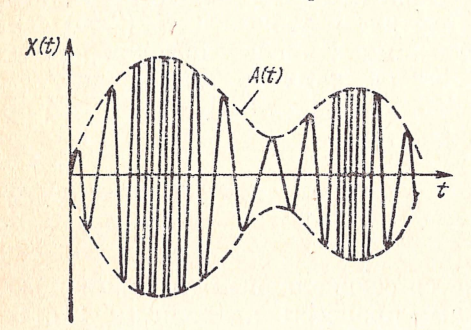


Рис. 2.5. Реализация узкополосного процесса

частотах. На экране осциллографа реализация узкополосного процесса имеет вид синусоиды с медленно меняющимися амплитудой и частотой (рис. 2.5).

Используют две основные равноценные формы аналитического представления узкополосных процессов: в виде амплитудно-частотно-модулированного колебания

$$X(t) = A(t)\cos[\omega_0 t + \Phi(t)], \tag{2.60}$$

где $A(t)$ — огибающая процесса, $\Phi(t)$ — фаза, и в виде суммы двух амплитудно-модулированных колебаний

$$X(t)=a(t)\cos\omega_0 t+b(t)\sin\omega_0 t, \qquad (2.61)$$

где

$$a(t)=A(t)\cos\Phi(t), \quad b(t)=A(t)\sin\Phi(t), \qquad (2.62)$$

$$A(t)=\sqrt{a^2(t)+b^2(t)}, \quad \Phi(t)=\mathrm{arctg}\,[b(t)/a(t)], \qquad (2.63)$$

Нетрудно заметить, что выбор формы связан с выбором системы координат. В полярной системе координат применяют представление (2.60), в декартовой — (2.61). Соотношения (2.62), (2.63) устанавливают связь между характеристиками узкополосного процесса в полярной и декартовой системе координат.

Представление (2.61) можно рассматривать и как частный случай ортогонального разложения (используется всего лишь одна гармоника). В то же время введение зависимости коэффициентов разложения от времени позволяет получить ряд полезных для описания модулированных сигналов свойств. Составляющую $a(t)$ называют синфазной, а $b(t)$ — квадратурной, говорят, что $a(t)$ и $b(t)$ находятся в квадратуре. Функции $a(t)$, $b(t)$, $A(t)$, $\Phi(t)$ являются медленно меняющимися функциями по отношению к гармоническому колебанию с частотой ω_0.

Функции $a(t)$ и $b(t)$ можно рассматривать и как ортогональные составляющие комплексной огибающей

$$\dot{A}(t)=a(t)+ib(t)=A(t)\,e^{i\Phi(t)}, \qquad (2.64)$$

а в более общем случае узкополосный процесс $X(t)$ — как вещественную часть комплексной функции

$$\dot{X}(t)=X(t)+i\hat{X}(t)=A(t)\exp\{i[\Phi(t)+\omega_0 t]\}, \qquad (2.65)$$

где

$$X(t)=\mathrm{Re}\,[\dot{X}(t)]=A(t)\cos[\Phi(t)+\omega_0 t],$$
$$\hat{X}(t)=I_m[\dot{X}(t)]=A(t)\sin[\Phi(t)+\omega_0 t]. \qquad (2.66)$$

Комплексная форма (2.65) записи узкополосного процесса является обобщением символической записи синусоидальных колебаний, в которой A и Φ рассматривают не как постоянные величины, а как функции времени.

Если $X(t)$ и $\hat{X}(t)$ составляют *пару преобразований Гильберта*

$$\hat{X}(t)=\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{X(\tau)}{t-\tau}d\tau, \quad X(t)=-\frac{1}{\pi}\int_{-\infty}^{\infty}\frac{\hat{X}(\tau)}{t-\tau}d\tau, \qquad (2.67)$$

то сигнал $\dot{X}(t)$ называют *аналитическим*.

Если сигнал $X(t)$ имеет непрерывный спектр

$$\Omega(i\omega) = \int_{-\infty}^{\infty} X(t)\, e^{-i\omega t}\, dt, \tag{2.68}$$

то спектр сопряженной функции $\hat{X}(t)$

$$\hat{\Omega}(i\omega) = [-i \operatorname{sign} \omega]\, \Omega(i\omega), \tag{2.69}$$

где знаковая функция

$$\operatorname{sign} \omega = \begin{cases} 1, & \omega > 0, \\ 0, & \omega = 0, \\ -1, & \omega < 0. \end{cases}$$

Следовательно, прямое преобразование Гильберта можно рассматривать как результат прохождения $X(t)$ через линейный четырехполюсник, сдвигающий фазу всех составляющих спектра на угол $-\pi/2$. Комплексная частотная и импульсная характеристики такого четырехполюсника:

$$\mathscr{K}(i\omega) = -i \operatorname{sign} \omega, \quad g(t) = 1/\pi t. \tag{2.70}$$

Спектр аналитического сигнала $\dot{X}(t)$

$$\Omega_X(i\omega) = \begin{cases} 2\Omega(i\omega), & \omega > 0, \\ \Omega(0), & \omega = 0, \\ 0 & \omega < 0. \end{cases} \tag{2.71}$$

Следовательно, спектр аналитического сигнала является односторонним и существует только в области положительных частот. Это удобное свойство.

Аналитические сигналы называют ортогональными в усиленном смысле, если справедливо условие

$$\frac{1}{T} \int_{-T/2}^{T/2} \dot{X}_i(t)\, \dot{X}^*{}_j(t)\, dt = 0, \quad i \neq j, \tag{2.72}$$

где звездочка обозначает величину, комплексно-сопряженную $\dot{X}_j(t)$.

Условие (2.72) равносильно совместному выполнению двух условий

$$\frac{1}{T} \int_{-T/2}^{T/2} X_i(t)\, X_j(t)\, dt = 0, \quad \frac{1}{T} \int_{-T/2}^{T/2} X_i(t)\, \hat{X}_j(t)\, dt = 0, \quad i \neq j.$$

Из соотношения (2.69) следует, что спектры и корреляционные функции случайных процессов $X(t)$ и $\hat{X}(t)$ одинаковы. Взаимный

энергетический спектр $\Omega_{12}(\omega)=i\Omega(\omega)$, а взаимнокорреляционная функция

$$K_{12}(\tau)=\frac{1}{\pi}\int\limits_0^\infty \Omega(\omega)\sin\omega\tau\, d\omega. \qquad (2.73)$$

Покажем, как определяется корреляционная функция узкополосного процесса. Рассмотрим процесс, спектральная плотность

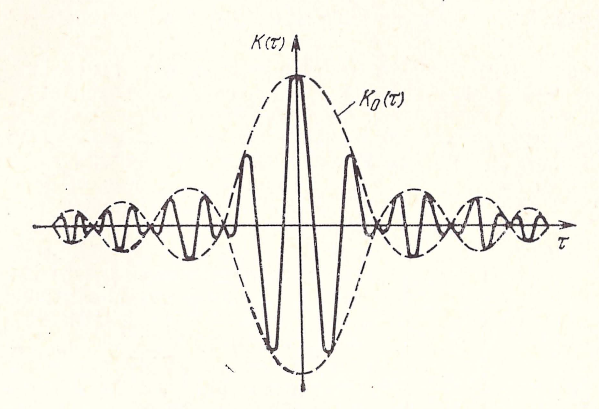


Рис. 2.6. Корреляционная функция узкополосного процесса

которого равномерна на интервале $[\omega_1, \omega_2]$ и для всех частот полосы $\Delta\omega=\omega_2-\omega_1$ равна Ω. Использовав (2.22), получим

$$K(\tau)=\frac{1}{\pi}\int\limits_{\omega_1}^{\omega_2}\Omega\cos\omega\tau\, d\omega=\frac{\Omega}{\pi\tau}(\sin\omega_2\tau-\sin\omega_1\tau)=$$

$$=\frac{\Delta\omega\Omega}{\pi}\;\frac{\sin(\Delta\omega\tau/2)}{\Delta\omega\tau/2}\cos\omega_0\tau=\mathscr{D}\rho_0(\tau)\cos\omega_0\tau, \qquad (2.74)$$

где $\mathscr{D}$ — дисперсия процесса; $\rho_0(\tau)=K_0(\tau)/\mathscr{D}_0$ — нормированная корреляционная функция огибающей; $K_0(\tau)$ — корреляционная функция огибающей.

На рис. 2.6 показан график корреляционной функции (2.74). Анализ (2.74) и рис. 2.6 позволяет сделать следующий общий вывод: для определения корреляционной функции узкополосного процесса необходимо найти корреляционную функцию огибающей и умножить ее на $\cos\omega_0\tau$.

Контрольные вопросы

1. Запишите и поясните условие узкополосности процесса.
2. Какие две основные формулы аналитического представления узкополосного процесса используют?
3. Как связаны огибающая и фаза с синфазной и квадратурной составляющими узкополосного процесса?

4. Как записывают узкополосный процесс в комплексной форме?
5. Какой сигнал называют аналитическим?
6. Запишите условие ортогональности сигналов в усиленном смысле.
7. Как определяют взаимный энергетический спектр и взаимную корреляционную функцию сигналов, сопряженных по Гильберту?
8. Как определить взаимный спектр и взаимную корреляционную функцию сигналов, сопряженных по Гильберту?
9. Как определить корреляционную функцию узкополосного процесса?

2.7. РАСПРЕДЕЛЕНИЯ ОГИБАЮЩЕЙ И ФАЗЫ УЗКОПОЛОСНЫХ СИГНАЛОВ

Распределения огибающей и фазы узкополосных сигналов получают как результат функционального преобразования системы случайных величин a, b в систему случайных величин A, Φ в соответствии с соотношениями (2.62), (2.63). Алгоритм решения этой задачи следующий. Определяют совместную плотность распределения f_{12} $(a,\ b)$ случайных величин a, b. Находят якобиан $\mathcal{J}$ преобразования от системы координат a, b к системе A, Φ. По плотности распределения $f_{12}(a,\ b)$, якобиану преобразования $\mathcal{J}$ и соотношениям (2.63) получают совместную плотность распределения $f_{21}(A,\ \Phi)$. Интегрированием этой плотности по «лишней» переменной находят одномерные плотности распределения огибающей $f_1(A)$ и фазы $f_2(\Phi)$.

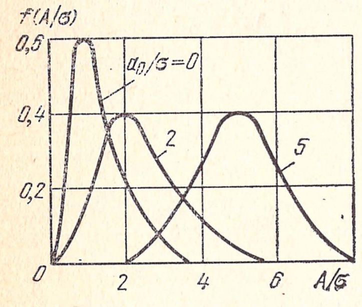


Рис. 2.7. Распределение Релея

Найдем по этому алгоритму законы распределения огибающей и фазы гауссовского стационарного узкополосного процесса. Предположим, что

$$M[X(t)]=0,\quad M[X^2(t)]=\mathcal{D},$$

$$K_{12}(\tau)=M[a(t)\,b(t+\tau)]=0.$$

Для гауссовского стационарного процесса квадратурные составляющие также будут нормальными стационарными процессами с параметрами

$$M[a(t)]=M[b(t)]=0,\quad M[a^2(t)]=M[b^2(t)]=\mathcal{D},$$

где $\mathcal{D}$ — дисперсия процесса; $K_{12}(\tau)$ — взаимная корреляционная функция квадратурных составляющих.

Совместная плотность распределения $a(t)$ и $b(t)$ в силу их некоррелированности равна произведению одномерных плотностей:

$$f_{12}(a,\ b)=\frac{1}{\sqrt{2\pi\mathcal{D}}}\exp\left(-\frac{a^2}{2\mathcal{D}}\right)\frac{1}{\sqrt{2\pi\mathcal{D}}}\exp\left(-\frac{b^2}{2\mathcal{D}}\right)=$$
$$=\frac{1}{2\pi\mathcal{D}}\exp\left(-\frac{a^2+b^2}{2\mathcal{D}}\right).\qquad(2.75)$$

Якобиан преобразования

$$\mathcal{J}=\begin{vmatrix}\partial a/\partial A & \partial a/\partial\Phi\\ \partial b/\partial A & \partial b/\partial\Phi\end{vmatrix}=\begin{vmatrix}\cos\Phi & -A\sin\Phi\\ \sin\Phi & A\cos\Phi\end{vmatrix}=A.\qquad(2.76)$$

Совместную плотность f_{21} $(A,\ \Phi)$ получим как произведение якобиана преобразования на совместную плотность распределения $f_{12}(a,\ b)$, в которой старые переменные a, b заменим на новые A, Φ, используя (2.63). Тогда

$$f_{2,1}(A,\ \Phi)=\frac{A}{2\pi\mathcal{D}}\,e^{-A^2/2\mathcal{D}}.\qquad(2.77)$$

Проинтегрировав (2.77) по фазе, получим одномерную плотность распределения огибающей

$$f_1(A) = \int_{-\pi}^{\pi} f_{21}(A, \Phi)\, dt = \frac{A}{\mathscr{D}} e^{-A^2/2\mathscr{D}}. \qquad (2.78)$$

Аналогично получим плотность распределения фазы

$$f_2(\Phi) = \int_{0}^{\infty} f_{21}(A, \Phi)\, dA = \frac{1}{2\pi}, \qquad (2.79)$$

так как

$$\int_{0}^{\infty} \frac{A}{\mathscr{D}} e^{-A^2/2D}\, dA = 1.$$

На рис. 2.7 показан график распределения (2.78), которое известно как распределение Релея, аргументом является безразмерная переменная — нормированная по среднеквадратическому отклонению огибающая A/σ $(\sigma=\sqrt{\mathscr{D}})$. Распределение огибающей существует на интервале $[0, \infty]$, максимальное значение плотность распределения имеет при $A/\sigma=1$.

Математическое ожидание огибающей

$$M[A] = \int_{0}^{\infty} A f_1(A)\, dA = \sqrt{\pi/2}\,\sigma \cong 1{,}25\sigma, \qquad (2.80)$$

дисперсия

$$\mathscr{D}[A] = \frac{4-\pi}{2}\sigma^2 \cong 0{,}43\mathscr{D}. \qquad (2.81)$$

Анализ (2.79) показывает, что распределение фазы является симметричным и равномерным. Поэтому математическое ожидание фазы

$$M[\Phi] = \int_{-\pi}^{\pi} \Phi f_2(\Phi)\, d\Phi = 0,$$

дисперсия фазы

$$\mathscr{D}[\Phi] = M[\Phi^2] = \pi^2/3. \qquad (2.82)$$

Контрольные вопросы

1. Изложите алгоритм определения плотностей распределений огибающей и фазы узкополосного процесса.
2. Проанализируйте свойства распределения Релея.
3. Проанализируйте свойства распределения фазы нормального узкополосного процесса.

2.8. РАСПРЕДЕЛЕНИЯ ОГИБАЮЩЕЙ И ФАЗЫ СУММЫ ГАРМОНИЧЕСКОГО КОЛЕБАНИЯ И УЗКОПОЛОСНОЙ ПОМЕХИ

Рассмотрим сигнал

$$z(t) = a_0 \cos\omega_0 t + \xi(t), \qquad (2.83)$$

который представляет сумму («смесь») гармонического колебания с амплитудой a_0 и узкополосной помехи $\xi(t)$. Найдем распределения огибающей и фазы $z(t)$.

Представим помеху в виде (2.61), тогда

$$z(t) = (a_0 + a_1)\cos\omega_0 t + b_1 \sin\omega_0 t, \qquad (2.84)$$

где a_1, b_1 — квадратурные составляющие помехи. Огибающая и фаза $z(t)$ соответственно записываются в виде

$$A = \sqrt{(a_0 + a_1)^2 + b_1^2}, \quad \Phi = \mathrm{arctg} \frac{b_1}{a_0 + a_1}. \tag{2.85}$$

Используя рассмотренный в § 2.7 алгоритм, находим

$$f_1(A) = \frac{A}{\sigma^2} I_0 \left(\frac{Aa_0}{\sigma^2}\right) \exp\left(-\frac{A^2 - a_0^2}{2\sigma^2}\right), \tag{2.86}$$

где σ^2 — дисперсия помехи; I_0 — модифицированная функция Бесселя нулевого порядка [1].

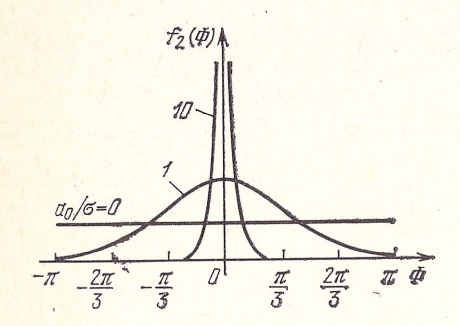


Рис. 2.8. Распределение фазы

Распределение (2.86) называют обобщенным релеевским распределением или распределением Райса. На рис. 2.7 показаны графики этого распределения при $a_0/\sigma=0$, 2, 5. Анализ графиков приводит к следующему *выводу*: при малых отношениях сигнал/шум ($a_0/\sigma \ll 1$) распределение (2.86) близко к распределению Релея; при больших отношениях сигнал/шум ($a_0/\sigma \gg 1$) оно стремится к нормальному распределению с математическим ожиданием, равным a_0, и дисперсией σ^2.

Плотность распределения фазы

$$f_2(\Phi) = \frac{1}{2\pi} \exp\left(-\frac{a_0^2}{2\sigma^2}\right) + \frac{a_0 \cos\Phi}{2\sqrt{2\pi\sigma^2}} \left[1 + F_{\text{К}}\left(\frac{a_0 \cos\Phi}{2\sigma^2}\right)\right] \times$$
$$\times \exp\left(-\frac{a_0^2 \sin^2\Phi}{2\sigma^2}\right), \tag{2.87}$$

где

$$F_{\text{К}}(u) = \frac{2}{\sqrt{2\pi}} \int_0^u e^{-z^2/2}\, dz. \tag{2.88}$$

— функция Крампа (интеграл вероятности) [2].

Функция Крампа связана с функцией Лапласа $\Phi_{\text{Л}}$ соотношением

$$F_{\text{К}}(u) = 2\Phi_{\text{Л}}(u) - 1, \tag{2.89}$$

где

$$\Phi_{\text{Л}}(u) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_0^u e^{-z^2/2}\, dz. \tag{2.90}$$

Функции Крампа и Лапласа табулированы [1—3] и, используя их, можно по величине u найти $F_{\text{К}}(u)$ [2]. Функция Крампа обладает следующими свойствами:

$\lim\limits_{u \to 0} F_К(u) = 0$, $\lim\limits_{u \to \infty} F_К(u) = 1$, $F_К(-u) = -F_К(u)$, т. е. функция Крампа является нечетной.

На рис. 2.8 приведены графики распределения (2.87) при $a_0/\sigma=0$, 1, 10. Анализ графиков показывает, что при малых отношениях сигнал/шум плотность распределения фазы близка к равномерной, при больших отношениях она стремится к дельта-функции в точке, соответствующей начальной фазе гармонического колебания.

Контрольные вопросы

1. Как ведет себя распределение огибающей смеси гармонического сигнала и узкополосной помехи при различных отношениях сигнал/шум?
2. Как ведет себя распределение фазы смеси гармонического сигнала и узкополосной помехи при различных отношениях сигнал/шум?

2.9. СИНТЕЗ СИГНАЛОВ И ПОМЕХ

Задачи синтеза сигналов и помех — это по существу задачи синтеза генераторов сигналов и помех. Задачи синтеза активных помех (в условиях радиопротиводействия) решают так же, как и задачи синтеза сигналов.

При создании переносчиков наметились две противоположные тенденции: создание высокостабильных узкополосных колебаний, ширину спектра которых стремятся сделать как можно меньше (приблизить к нулю), и создание широкополосных (шумоподобных) сигналов, ширину спектра которых стремятся увеличить как можно больше. Это объясняется тем, что использование и тех и других переносчиков позволяет получить определенные преимущества. Рассмотрим современные принципы генерации и особенности применения узкополосных и широкополосных переносчиков.

Для генерации высокостабильных узкополосных колебаний все большее распространение получают аналоговые и цифровые синтезаторы частот. Они позволяют получить десятки тысяч высокостабильных колебаний в заданном диапазоне частот. Задачи синтеза таких сигналов сводятся по существу к задачам оптимального построения синтезаторов частот, обладающих требуемыми технико-экономическими характеристиками: рабочим диапазоном частот, стабильностью колебаний, гибкостью и оперативностью управления частотой сигналов, технологичностью, габаритами, массой, потребляемой мощностью, приведенными годовыми расходами, оптовой ценой, эксплуатационными расходами и т. п.

Синтезаторы частот широко применяют в технике связи, информационно-измерительной технике и других областях. Способы построения синтезаторов детально рассматривают в курсе «Радиопередающие устройства». Для иллюстрации особенностей синтеза сигналов мы остановимся лишь на одном цифровом методе формирования высокостабильной сетки частот. Он был предложен в 1966—1967 гг. зарубежными специалистами З. Блаховицем, А. Эверсом и советскими учеными О. Губернаторовым и В. Пивоваром. Сущность цифрового принципа формирования и стабилизации дискретного множества частот заключается в использовании свойств системы фазовой автоподстройки частоты с делителем частоты в цепи обратной связи и с предварительным преобразованием гармонических колебаний управляемого и опорного генераторов с помощью формирующих устройств в последовательность видеоимпульсов. Это позволяет строить синтезаторы на элементах вычислительной техники и называть их цифровыми.

На рис. 2.9 приведена структурная схема цифрового синтезатора частот. Работу схемы поясняют временные диаграммы напряжений на выходе каждого устройства. В схеме имеются два генератора синусоидальных колебаний: управляемый генератор, период колебаний которого T_1, и опорный генератор, стабилизированный кварцем, период колебаний которого $T_0=kT_1$.

В формирующих устройствах гармонические колебания преобразуются в последовательности видеоимпульсов с периодами соответственно T_1 и T_0. Делитель частоты с переменным коэффициентом деления k преобразует последователь-

ность видеоимпульсов с периодом T_1 в последовательность с периодом kT_1. В фазоимпульсном детекторе фазы импульсов полученных последовательностей сравниваются и на выходе фильтра нижних частот появляется пилообразное напряжение как результат рассогласования фаз. Это напряжение через управляющий элемент так воздействует на управляемый генератор, чтобы в установившемся режиме соблюдалось равенство $T_0=kT_1$.

Основные достоинства цифрового метода синтеза сигналов следующие: генерация сколь угодно густой сетки высокостабильных колебаний, возможность использования унифицированных элементов вычислительной техники, выполнен-

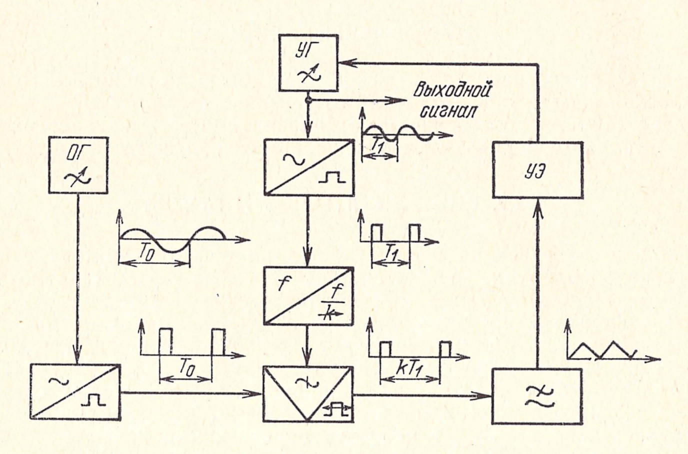


Рис. 2.9 Структурная схема цифрового синтезатора частот

ных на полупроводниковых интегральных и гибридных микросхемах, и применения прогрессивной технологии сборки и настройки, малые габариты, масса и затраты энергии и т. п.

Для генерации широкополосных (шумоподобных) сигналов-переносчиков, свойства которых близки к свойствам нормального случайного процесса, используют шумовые и псевдошумовые генераторы (генераторы детерминированных сигналов, построенных по определенным законам). Работа таких генераторов основана на следующих принципах: сложение сигналов на выходе параллельно соединенных мультивибраторов, амплитуды импульсов которых выбирают так, чтобы получить требуемые корреляционные и спектральные характеристики суммарного шумоподобного сигнала; синтез на основе ортогонального разложения Котельникова; модуляция высокостабильного колебания шумоподобным видеосигналом; периодическое повторение отрезка высокочастотного радиосигнала, обладающего требуемыми корреляционными и спектральными характеристиками, и др.

На рис. 2.10 представлена упрощенная структурная схема генератора широкополосных сигналов, которая создает сигнал на основе ортогонального разложения Котельникова. При подаче с генератора на вход узкополосных фильтров импульса с малой длительностью каждый из фильтров «вырезает» из непрерывного спектра этого импульса узкую область частот в соответствии с полосой прозрачности. Ослабление k-го сигнала в k-м аттенюаторе в β_k раз и задержка его на время τ_k позволяют получить на выходе сумматора шумоподобный сигнал

$$s(t)=\sum_{k=1}^{n}\beta_k\frac{\sin[\Delta\omega(t-\tau_k)/2]}{\Delta\omega(t-\tau_k)/2}\,e^{i\omega_k(t-\tau_k)}. \qquad (2.91)$$

Каждая составляющая суммы представляет собой узкополосный процесс в комплексной форме (2.65), роль огибающей играет произведение при показательной функции. Фильтры имеют полосу $\Delta\omega$ и центральную частоту ω_k.

Рассмотренная схема позволяет синтезировать непрерывные сигналы с ограниченным спектром любого типа, в том числе и шумоподобные. Управляемыми переменными служат величины ω_k, $\Delta\omega$, β_k, τ_k, n. Для получения ансамбля таких сигналов достаточно изменять β_k и τ_k. При больших базах генерируемых сигналов схема становится громоздкой. Ее можно упростить, применяя общие линию задержки и амплитудный корректор. Достоинством этого способа формирования сигналов является возможность построения согласованных фильтров для оптимального приема таких сигналов. Согласованные фильтры составляются из тех же элементов, но с обратным порядком включения.

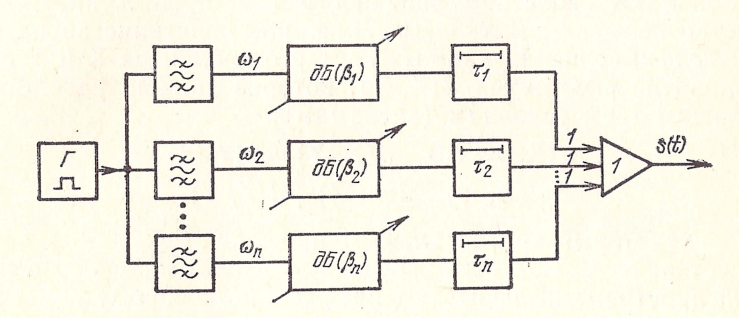


Рис. 2.10. Структурная схема генератора широкополосных сигналов

Применение шумоподобных сигналов позволяет: принимать сигналы и в тех случаях, когда отношение сигнал/шум гораздо меньше единицы; бороться с вредным влиянием многолучевого распространения радиоволн; ослабить влияние на передачу информации узкополосных помех; обеспечить высокий коэффициент использования пропускной способности каналов; использовать модуляцию сигналов по форме и т. п.

При решении задач синтеза сигналов с общих позиций целесообразно учитывать не только структуру и параметры сигналов, но и технико-экономические характеристики генераторов сигналов, свойства модулирующих сигналов, вид модуляции и характер последующих преобразований модулированных колебаний. Эти задачи находятся в стадии научных исследований.

Контрольные вопросы

1. Какие тенденции наблюдаются в использовании сигналов-переносчиков?
2. В чем сущность задачи синтеза сигналов?
3. Поясните цифровой способ формирования высокостабильной сетки частот.
4. Какие способы используют при генерации широкополосных сигналов?
5. Поясните способ синтеза широкополосного сигнала на основе ортогонального разложения Котельникова.

2.10. ПРОСТРАНСТВА СИГНАЛОВ И ПОМЕХ

При решении задач теории информации и передачи сигналов используют также векторное представление сигналов и помех, методы аналитической геометрии и векторной алгебры.

Рассмотрим особенности векторного представления сигналов и помех как элементов функциональных пространств.

Сигналы и помехи рассматривают как векторы — элементы векторных функциональных пространств, а преобразования сигналов

и помех — как отображения одних пространств в другие. Наряду с рассмотренными в приложении понятиями функционального анализа применяют и такие понятия, как метрика пространства, расстояние, норма вектора, проекция вектора и др. Наибольшее распространение получило использование пространств Евклида, Гильберта и Хемминга.

Если преобразования элементов в функциональном пространстве обладают свойством линейности (см. приложение), то пространство называют *линейным*. Линейное пространство называют *метрическим*, если для каждых двух его элементов $\mathbf{X}$ и $\mathbf{Y}$ определено понятие расстояния $d(\mathbf{X}, \mathbf{Y})$, которое должно удовлетворять следующим трем условиям (аксиомам):

$$d(\mathbf{X}, \mathbf{X})=0, \quad d(\mathbf{X}, \mathbf{Y})=d(\mathbf{Y}, \mathbf{X}),$$
$$d(\mathbf{X}, \mathbf{Y}) \leqslant d(\mathbf{X}, \mathbf{Z})+d(\mathbf{Z}, \mathbf{Y}), \qquad (2.92)$$

где $\mathbf{Z}$ — некоторый третий элемент пространства.

Расстояние может быть введено различными способами. *Метрикой* пространства называют правило, по которому введено расстояние. В рассматриваемых пространствах метрика определяется через операцию скалярного произведения векторов. Скалярное произведение двух векторов в пространстве Евклида — это число:

$$\mathbf{XY}=\sum_{k=1}^{n} x_k y_k, \qquad (2.93)$$

где x_k, y_k — координаты векторов $\mathbf{X}$, $\mathbf{Y}$; n — размерность пространства. При $n=3$ пространство Евклида является математическим прообразом реального трехмерного пространства. *Пространство Евклида* обозначают обычно R_2.

Скалярное произведение двух векторов в *пространстве Гильберта* (пространство всех непрерывных функций времени, заданных на интервале $[0, T]$) — это число

$$\mathbf{XY}=\frac{1}{T}\int_{0}^{T} x(t)\, y(t)\, dt. \qquad (2.94)$$

Пространство Гильберта обозначают обычно L_2.

Норму l_x *вектора* $\mathbf{X}$ в пространстве R_2 определяют через скалярное произведение следующим образом:

$$l_x=\sqrt{\mathbf{XX}}=\sqrt{\sum_{k=1}^{n} x^2_k}. \qquad (2.95)$$

Следовательно, норма вектора — это его длина. Аналогично определяют *норму с вектора* $\mathbf{X}$ в пространстве L_2:

$$c=\sqrt{\mathbf{XX}}=\sqrt{\frac{1}{T}\int_{0}^{T} x^2(t)\, dt}. \qquad (2.96)$$

Если сигнал $x(t)$ имеет размерность напряжения, то норма — это эффективное значение напряжения, а квадрат нормы — это средняя мощность сигнала.

Если непрерывный сигнал с ограниченным спектром представлен в виде разложения Котельникова, то энергия сигнала (см. (2.7))

$$E=\int_0^T x^2(t)\,dt=\frac{1}{2F}\sum_{k=1}^{n} x^2{}_k.$$

Так как средняя мощность сигнала $P=c^2=E/T$, то $E=PT$ и

$$\sum_{k=1}^{n} x^2{}_k=2FTP=nP. \qquad (2.97)$$

Из (2.97) нетрудно установить связь между нормами векторов в пространствах R_2 и L_2:

$$l^2{}_x=nc^2, \quad c=l_x/\sqrt{n}. \qquad (2.98)$$

Расстояние между векторами **X** и **Y** в пространстве R_2 определяют как норму их разности

$$d(\mathbf{X},\mathbf{Y})=\sqrt{\sum_{k=1}^{n}(x_k-y_k)^2}. \qquad (2.99)$$

В пространстве L_2 расстояние определяют аналогично

$$d(\mathbf{X},\mathbf{Y})=\sqrt{\frac{1}{T}\int_0^T [x(t)-y(t)]^2\,dt}. \qquad (2.100)$$

Использование расстояния позволяет определять *проекции одного вектора на другой*. Рассмотрим два процесса $X(t)$ и $Y(t)$, которые имеют одинаковую длительность T и одинаковую граничную частоту F. Возьмем $n=2TF$ отсчетов $X(t)$ и $Y(t)$ и образуем вектора **X** и **Y** с координатами x_k и y_k.

Квадрат расстояния между этими векторами в соответствии с (2.99) равен

$$d^2(\mathbf{X},\mathbf{Y})=\sum_{k=1}^{n} x^2{}_k+\sum_{k=1}^{n} y^2{}_k-2\sum_{k=1}^{n} x_k y_k.$$

Угол θ между векторами **X** и **Y** определяют из выражения

$$\cos\theta=\mathbf{XY}/l_x l_y,$$

поэтому

$$d^2(\mathbf{X},\mathbf{Y})=l^2{}_x+l^2{}_y-2l_x l_y\cos\theta. \qquad (2.101)$$

С учетом (2.97) и (2.99) запишем (2.101) в более удобном для анализа виде

$$d^2(\mathbf{X}, \mathbf{Y}) = 2TF\,[P_1 + P_2 - 2\sqrt{P_1P_2}\cos\theta], \qquad (2.102)$$

где P_1, P_2 — средняя мощность процессов $X(t)$ и $Y(t)$.

Анализ (2.102) показывает, что расстояние между сигналами определяется их базой, мощностью сигналов и значением $\cos\theta$. Если $X(t)$ и $Y(t)$ — случайные процессы, то $\cos\theta$ играет роль коэффициента корреляции векторов **X** и **Y**.

Если $\cos\theta = 0$ $(\theta = \pi/2)$, *векторы не коррелированы*, т. е. они ортогональны, и проекция одного вектора на другой равна нулю. Если $\cos\theta = 1$ $(\theta = 0)$, то **X** и **Y** полностью коррелированы (совпадают по направлению). Если $\cos\theta = -1$ $(\theta = \pi)$, то **X** и **Y** также полностью коррелированы, но эти векторы противоположно направлены. Такие сигналы называют *противоположными*, расстояние между ними наибольшее. Расстояние между сигналами играет существенную роль при разделении сигналов в многоканальных системах, при выделении полезных сигналов и в других задачах.

Рассмотрим связь элементов пространств R_2 и L_2. Если в пространстве R_2 с координатами векторов $x_k/\sqrt{n}$, $y_k/\sqrt{n}$ размерность пространства $n \to \infty$, то евклидово пространство переходит в гильбертово. Напомним (см. (2.7)—(2.9), (2.14)), что в этом случае ортогональное разложение Котельникова непрерывного сигнала в дискретную последовательность отсчетов дает нулевую среднеквадратическую погрешность. Отсчеты непрерывной функции пространства L_2 играют роль координат вектора в пространстве R_2. Ортогональное разложение Котельникова можно рассматривать как скалярное произведение двух векторов: вектора отсчетов и базисного вектора. Тогда это разложение устанавливает связь между элементами гильбертова и евклидова пространства.

При преобразованиях сигналов рассматривают пространства полезных сигналов (сообщений), пространства модулированных сигналов, пространства принятых сигналов, пространства демодулированных сигналов и др. Последовательные преобразования сигналов рассматривают как взаимные отображения одних пространств в другие, устанавливающие соответствия между элементами различных пространств. Например, если сообщение (видеосигнал) $u(t)$ является непрерывным сигналом с ограниченным спектром, то он может быть представлен вектором в n_1-мерном пространстве, где $n_1 = 2T_1F_1$; T_1, F_1 — длительность и граничная частота спектра. При модуляции сигнал $u(t)$ преобразуется в модулированный сигнал $s(t)$, следовательно, пространство сообщений преобразуется в пространство модулированных сигналов. Если длительность $s(t)$ равна T_2, а граничная частота F_2, то размерность нового пространства $n_2 = 2T_2F_2$. В общем случае $n_2 > n_1$ и только при однополосной модуляции $n_2 = n_1$.

Пространство Хемминга (пространство двоичных сигналов) относится к линейным дискретным пространствам, особенностью ко-

торых является то, что координаты векторов могут принимать лишь дискретные значения. Понятия скалярного произведения и нормы для таких пространств вводят подобно рассмотренным. Метрику и расстояние вводят на основе операции сложения по модулю одноименных разрядов (см. § 1.6, (1.18), (1.19)).

Контрольные вопросы

1. Какое пространство называют линейным и какое метрическим?
2. Каким условиям должно удовлетворять расстояние в функциональном пространстве?
3. Как определяют скалярное произведение в пространствах R_2, L_2?
4. Как определяют норму вектора в пространствах R_2, L_2?
5. Какой физический смысл имеет норма сигнала?
6. Как связаны нормы сигнала в пространствах R_2, L_2?
7. Как определяют расстояние в пространствах R_2, L_2?
8. Чем определяется расстояние между двумя сигналами?
9. Какие сигналы называют противоположными?
10. Какая связь между элементами пространств R_2, L_2?
11. Какие пространства рассматривают при преобразованиях сигналов?
12. Какая основная особенность линейных дискретных пространств?

2.11. ВЫВОДЫ

1. Обобщенной спектральной теорией сигналов называют совокупность методов аналитического представления сигналов в виде (1.3). Это одна из наиболее удобных форм описания сигналов для анализа линейных и нелинейных систем. Обобщенная спектральная теория исследует основные закономерности спектрального анализа, общие для различных систем базисных функций, и ставит задачи оптимального выбора этих систем для успешного решения задач передачи и обработки сигналов.

2. Ортогональные разложения Котельникова (2.13), (2.15) для непрерывных сигналов с ограниченными и полосовыми спектрами, так же как и преобразования Фурье для периодических и непериодических сигналов, являются практически важными частными случаями обобщенного ряда Фурье (2.5), примерами практического применения обобщенной спектральной теории.

Ряды Котельникова позволяют представить непрерывные сигналы в виде дискретных последовательностей импульсов, отстоящих друг от друга на интервал дискретизации. Этот интервал полностью определяется верхней граничной частотой для сигналов с ограниченным спектром и шириной спектра для сигналов с полосовыми спектрами.

Наиболее важным для практики свойством рядов Котельникова является относительно простая аппаратурная реализация дискретизации и восстановления непрерывных сигналов. Поэтому ортогональные разложения Котельникова служат основой для построения дискретных методов передачи непрерывных сигналов. Во многих случаях они позволяют с единых позиций рассматривать передачу дискретных и непрерывных сигналов.

3. При решении задач теории информации и передачи сигналов широко используют такие характеристики сигналов, как корреляционная функция, спектральная плотность распределения мощности, дисперсия, интервал корреляции, ширина спектра, взаимокорреляционная функция двух сигналов и взаимная спектральная плотность распределения мощности двух процессов.

Корреляционная функция показывает характер статистической связи двух значений случайного процесса, отстоящих друг от друга на некоторый интервал времени. Для определения корреляционных функций процессов используют операции усреднения по множеству (2.17) и по времени (2.20). Связь между корреляционными и спектральными характеристиками случайного процесса устанавливают преобразования Хинчина — Винера (2.22), которые являются аналогом преобразований Фурье для детерминированного процесса.

4. Для моделирования случайных сигналов и помех в теории информации и передачи сигналов часто используют телеграфный сигнал—случайный процесс с показательной корреляционной функцией (2.27), белый шум, гауссовский процесс и гауссовский белый шум. Случайный процесс с показательной корреляционной функцией обладает полезной особенностью. Изменением единственного параметра α (2.27) можно в широких пределах изменять корреляционные и спектральные свойства процесса. При $\alpha \to 0$ процесс вырождается в детерминированный, при $\alpha \to \infty$ — в белый шум.

Белый шум является стационарным случайным процессом с постоянной спектральной плотностью мощности на всех частотах, его используют как модель наиболее тяжелого вида помех. Гауссовским называют процесс, который имеет нормальное распределение мгновенных значений (2.40). Нормально распределенное колебание образуется в результате сложения большого числа независимых или слабокоррелированных случайных колебаний. Белый шум, у которого распределение мгновенных значений является нормальным, называют гауссовским белым шумом.

5. Канонические (2.44) и неканонические (2.55) разложения случайных сигналов и помех удобны для решения многих задач. Для стационарных случайных процессов часто используют тригонометрические ряды Фурье (2.45), в которых коэффициенты разложения являются некоррелированными случайными величинами. Для нестационарных процессов необходимо выбирать другие базисные функции, чтобы обеспечить некоррелированность коэффициентов разложения.

В каноническом разложении Котельникова интервал дискретизации случайного процесса определяется его интервалом корреляции, максимальным значением спектральной плотности и значением спектральной плотности на нулевой частоте. Интервал дискретизации больше или равен интервалу корреляции процесса.

6. Как специальный вид ортогональных разложений можно рассматривать представление реальных сигналов и помех в виде узкополосных (2.60), (2.61) и аналитических сигналов (2.65).

Особенность такого представления проявляется в том, что коэффициенты разложения являются функциями времени. Процесс называют узкополосным, если ширина спектра процесса относительно мала по сравнению со средней частотой спектра (2.59). Понятие об аналитических сигналах основано на обобщении символической записи синусоидальных колебаний в комплексной форме. Спектр аналитического сигнала существует только в области положительных частот. Корреляционная функция узкополосного процесса (2.74) равна произведению корреляционной функции огибающей (см. (2.60)) на $\cos \omega_0\tau$, где ω_0 — средняя частота спектра процесса.

7. Распределение огибающей гауссовского стационарного узкополосного процесса является релеевским распределением (2.78), распределение фазы — равномерным (2.79).

8. Распределение огибающей суммы гармонического колебания и гауссовского стационарного узкополосного процесса является обобщенным распределением Релея (распределением Райса) (2.86), распределение (2.87) фазы суммы определяется через функцию Крампа (2.88).

9. Задачи синтеза сигналов и помех сводятся к задачам разработки генераторов сигналов и помех. Для получения сигналов-переносчиков создают: высокостабильные узкополосные колебания с возможно меньшей шириной спектра и широкополосные (шумоподобные) колебания с возможно большей шириной спектра. Применение и тех и других переносчиков имеет свои преимущества. При решении задач синтеза сигналов в общей постановке необходимо учитывать не только структуру и параметры сигналов, но и характеристики генераторов, свойства полезных сигналов, вид модуляции, а также характер последующих преобразований модулированных сигналов.

10. Векторное представление сигналов и помех позволяет применять для решения задач теории информации и передачи сигналов известные методы аналитической геометрии, векторной алгебры и функционального анализа. Сигналы и помехи рассматривают как элементы функциональных пространств, а преобразования сигналов и помех — как отображения одних пространств в другие. Наибольшее распространение получили пространства Евклида, Гильберта и Хемминга.

Глава 3

УПРАВЛЕНИЕ ИНФОРМАЦИОННЫМИ ПАРАМЕТРАМИ СИГНАЛОВ

3.1 КЛАССИФИКАЦИЯ МЕТОДОВ МОДУЛЯЦИИ

Исследование различных видов модуляции необходимо для определения требуемых свойств каналов, сокращения избыточности модулированных сигналов и улучшения использования мощности передатчиков, определения потенциальной помехоустойчивости, помех соседним каналам и успешного решения проблем электромагнитной совместимости радиосистем; разработки оптимальных методов аппаратурной реализации модуляции.

Идеальная непрерывная модуляция — это перенос спектра полезного сигнала в область более высоких частот без нелинейных, частотных и фазовых искажений. Если полезный сигнал пердставить в виде узкополосного процесса

$$u(t)=A(t)\cos[\omega_0 t+\Phi(t)], \tag{3.1}$$

то модулированный сигнал в идеальном случае должен иметь вид

$$s(t)=A(t)\cos[\omega_1 t+\Phi(t)], \tag{3.2}$$

где $\omega_1=\omega_0+\omega_2$; ω_2 — средняя частота сигнала-переносчика.

Из соотношений (3.1) и (3.2) следует, что при идеальной модуляции законы распределения огибающей и фазы узкополосного сигнала не должны изменяться, изменяется только средняя частота. Корреляционная функция огибающей не изменяется, частота «косинусоидального заполнения» корреляционной функции модулированного сигнала равна ω_1. Спектр модулированного сигнала смещается в область средней частоты ω_1, но не изменяет своей формы.

Реально модуляция сопровождается нелинейными, частотными и фазовыми искажениями. Поэтому, как правило, ширина спектра модулированных сигналов, больше ширины спектра полезных сигналов, искажаются законы распределения огибающей и фазы, изменяются формы корреляционных функций и спектральных плотностей.

Если полезный сигнал является случайным стационарным процессом, а сигнал-переносчик — гармоническим колебанием, то модулированный сигнал уже является нестационарным случайным процессом, корреляционная функция и спектральная плотность которого зависят от текущего момента времени.

Для определения средних спектральных и корреляционных характеристик модулированного сигнала необходимо дополнительно применять операцию усреднения по времени. В этом более общем случае соотношения Хинчина — Винера принимают вид

$$\Omega_1(\omega) = 2\int_0^\infty \langle K(\tau, t)\rangle \cos\omega\tau\, d\tau, \tag{3.3}$$

$$K_1(\tau) = \frac{1}{\pi}\int_0^\infty \langle \Omega(\omega, t)\rangle \cos\omega\tau\, d\omega, \tag{3.4}$$

где $\Omega_1(\omega)$, $K_1(\tau)$ — средняя спектральная плотность и средняя корреляционная функция модулированного сигнала.

Корреляционные функции, спектральные плотности, законы распределения огибающей и фазы модулированных сигналов получают по заданным операторам модуляции, корреляционным функциям или спектральным плотностям, законам распределения огибающих и фаз полезного сигнала и переносчика.

Таблица 1

Классы $x(t)$	Классы $u(t)$							
	A	B	C	D	E	F	G	H
1	A1	B1	C1	D1	E1	F1	G1	H1
2	A2	B2	C2	D2	E2	F2	G2	H2
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
8	A8	B8	C8	D8	E8	F8	G8	H8

Для классификации видов модуляции удобно использовать следующие признаки: характер полезного сигнала и переносчика (детерминированный процесс, случайный стационарный процесс, случайный нестационарный процесс); вид сигналов (аналоговые, дискретные); вид информационного параметра (амплитуда, частота, фаза, форма, длительность, период и т. п.) и др.

Не затрагивая прикладной классификации типов модуляции по МККТТ (Международный консультативный комитет по телефонии и телеграфии), которая излагается в специальных курсах, рассмотрим построение классификационной матрицы (табл. 1) видов мо-

дуляции в том случае, когда учитывают два основных признака: вид модулирующего сигнала и вид переносчика. Условно введем следующие классы модулирующих сигналов $u(t)$: класс **A** — детерминированные непрерывнозначные процессы, класс **B** — детерминированные дискретные последовательности, класс **C** — случайные стационарные непрерывнозначные процессы, класс **D** — случайные стационарные последовательности, класс **E** — случайные нестационарные непрерывнозначные процессы, класс **F** — случайные нестационарные последовательности, класс **G** — дискретные случайные стационарные последовательности, класс **H** — дискретные случайные нестационарные последовательности (см. ГОСТ 21878—76 «Случайные процессы и динамические системы»). Аналогично введем классы переносчиков $x(t)$, то для удобства записи обозначим их соответственно цифрами 1—8. В соответствии с введенными обозначениями класс **A**1 включает все непрерывные виды модуляции, в которых полезные сигналы и переносчики являются детерминированными непрерывными процессами; класс **B**1 — все виды модуляции, в которых полезный сигнал рассматривают как детерминированную импульсную последовательность, а переносчик — как детерминированный непрерывный сигнал.

Аналогично объединяют в классы остальные виды модуляции. Детализацию и углубление этой классификации можно выполнить, если учесть и другие признаки полезных сигналов и переносчиков. Если последовательно выполняют несколько различных методов модуляции (применяют многоступенчатую модуляцию), получившиеся смешанные классы обозначают **C**1**C**3, **A**1**A**2 и т. д. В настоящее время наиболее изучены простейшие виды модуляции классов **A**1, **A**2, **B**1, **B**2, исследуют различные виды модуляции классов **C**1, **C**2, **D**1, **D**2. В связи с развитием лазерной и космической техники, спутниковой связи большое внимание уделяют классам **C**3, **C**4, **D**3 и **D**4. Внедрение в технику связи цифровых способов передачи, развитие сетей связи, исследование возможностей создания интегральных сетей связи, в которых передачу сигналов и управление информационными потоками выполняют в цифровой форме, порождают необходимость изучения классов **G**1—**G**4, **H**1—**H**4.

В теории информации и передачи сигналов основное внимание уделяют тем классам модуляции, в которых полезные сигналы рассматривают как случайные. Это обусловлено тем, что детерминированные сигналы не несут информации. Далее рассматриваются корреляционные и спектральные характеристики модулированных случайных сигналов, анализируются характеристики модулированных сигналов классов **C**3, **A**2, **B**1, **C**2, **C**3, **D**1, излагается сущность цифровых методов модуляции классов **G**1—**G**3.

Контрольные вопросы

1. Для чего необходимо изучать различные виды модуляции?
2. В чем сущность идеальной непрерывной модуляции?
3. Какие особенности определения корреляционных и спектральных характеристик модулированных сигналов?

4. Что необходимо знать для анализа характеристик модулированных сигналов?
5. Укажите признаки классификации видов модуляции.
6. Как строится классификационная матрица видов модуляции?

3.2. КОРРЕЛЯЦИОННЫЕ И СПЕКТРАЛЬНЫЕ ХАРАКТЕРИСТИКИ МОДУЛИРОВАННЫХ СИГНАЛОВ

Особенности определения корреляционных и спектральных характеристик модулированных сигналов рассмотрим для непрерывных видов модуляции. Для других видов модуляции эти характеристики изучают аналогично. Для многих практически важных видов непрерывной модуляции модулированный сигнал можно рассматривать как узкополосный процесс в виде (3.2). Поэтому характеристики модулированного сигнала изучают методами, изложенными в § 2.6. Покажем, как определяют корреляционную функцию и спектральную плотность модулированного сигнала.

Для определения корреляционной функции применим операцию усреднения по времени:

$$K(\tau) = \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} s(t)\, s(t-\tau)\, dt \approx \int_{-\infty}^{\infty} s(t)\, s(t-\tau)\, dt.$$

Выразим $\cos[\omega_1 t + \Phi(t)]$ в (3.2) через показательные функции по формуле Эйлера, тогда

$$K(\tau) = \int_{-\infty}^{\infty} A(t) \frac{e^{i[\omega_1 t + \Phi(t)]} + e^{-i[\omega_1 t + \Phi(t)]}}{2} \times$$

$$\times A(t-\tau) \frac{e^{i[\omega_1(t-\tau) + \Phi(t-\tau)]} + e^{-i[\omega_1(t-\tau) + \Phi(t-\tau)]}}{2}\, dt.$$

После перемножения функций, стоящих под интегралом, получим

$$K(\tau) = \frac{1}{4} \left\{ \int_{-\infty}^{\infty} A(t) A(t-\tau) \exp\{i[\omega_1\tau + \Phi(t) + \Phi(t-\tau)]\} \exp(2i\omega_1 t)\, dt + \right.$$

$$+ \int_{-\infty}^{\infty} A(t) A(t-\tau) \exp\{-i[\omega_1\tau + \Phi(t) - \Phi(t-\tau)]\} \exp(-2i\omega_1 t)\, dt +$$

$$+ \int_{-\infty}^{\infty} A(t) A(t-\tau) [\exp\{i[\omega_1\tau + \Phi(t-\tau) - \Phi(t)]\} +$$

$$\left. + \exp\{-i[\omega_1\tau + \Phi(t-\tau) - \Phi(t)]\}]\, dt \right\}.$$

В первых двух интегралах множители $\exp(2i\omega_1 t)$ и $\exp(-2i\omega_1 t)$ являются быстро изменяющимися по сравнению с функциями $A(t)$, $A(t-\tau)$, $\Phi(t)$ и $\Phi(t-\tau)$, поэтому значениями этих интегралов можно пренебречь по сравнению со значением третьего интеграла. Следовательно,

$$K(\tau) = \frac{1}{2} \int_{-\infty}^{\infty} A(t) A(t-\tau) \cos[\omega_1\tau + \Phi(t-\tau) - \Phi(t)]\, dt. \tag{3.5}$$

Выражение (3.5) является основным для определения корреляционных функций модулированных сигналов при различных видах непрерывной модуляции.

Рассмотрим для примера балансную модуляцию случайным процессом $u(t)$ с корреляционной функцией $K_1(\tau)=\mathscr{D}e^{-\alpha|\tau|}$ гармонического колебания с единичной амплитудой $x(t)=\cos\omega_0 t$. В этом случае $s(t)=u(t)\cos\omega_0 t$ и интеграл (3.5) принимает простой вид:

$$K(\tau)=\frac{1}{2}\cos\omega_0\tau\int\limits_{-\infty}^{\infty}u(t)\,u(t-\tau)\,dt.$$

Интеграл в правой части является корреляционной функцией $K_1(\tau)$ огибающей $u(t)$, а $(\cos\omega_0\tau)/2$ — корреляционной функцией $K_2(\tau)$ гармонического колебания. Поэтому, как и следовало ожидать (см. § 2.6), корреляционная функция гармонического колебания, балансно-модулированного случайным процессом,

$$K(\tau)=K_1(\tau)\,K_2(\tau)=0{,}5\mathscr{D}e^{-\alpha|\tau|}\cos\omega_0\tau. \tag{3.6}$$

Спектральную плотность гармонического колебания, балансно-модулированного случайным процессом, определим с помощью соотношения Хинчина — Винера (2.22):

$$\Omega(\omega)=\int\limits_{-\infty}^{\infty}\frac{\mathscr{D}}{2}e^{-\alpha|\tau|}\cos\omega_0\tau e^{-i\omega\tau}d\tau=\frac{\mathscr{D}}{2}\int\limits_{-\infty}^{\infty}e^{-\alpha|\tau|}\frac{e^{i\omega_0\tau}+e^{-i\omega_0\tau}}{2}e^{-i\omega\tau}d\tau=$$

$$=\frac{\mathscr{D}}{4}\left\{\int\limits_{-\infty}^{0}e^{\tau[i(\omega_0-\omega)+\alpha]}d\tau+\int\limits_{0}^{\infty}e^{\tau[i(\omega_0-\omega)-\alpha]}d\tau+\right.$$

$$\left.+\int\limits_{-\infty}^{0}e^{\tau[-i(\omega_0+\omega)+\alpha]}d\tau+\int\limits_{0}^{\infty}e^{\tau[-i(\omega_0+\omega)-\alpha]}d\tau\right\}=\frac{\mathscr{D}}{4}\left[\frac{1}{i(\omega_0-\omega)+\alpha}-\right.$$

$$\left.-\frac{1}{i(\omega_0-\omega)-\alpha}+\frac{1}{-i(\omega_0+\omega)+\alpha}-\frac{1}{-i(\omega_0+\omega)-\alpha}\right].$$

Окончательно имеем

$$\Omega(\omega)=\frac{\alpha\mathscr{D}}{2}\left[\frac{1}{(\omega_0+\omega)^2+\alpha^2}+\frac{1}{(\omega_0-\omega)^2+\alpha^2}\right]. \tag{3.7}$$

Следовательно, спектр гармонического колебания, балансно-модулированного случайным процессом, имеет две боковые полосы частот в области ω_0 и $-\omega_0$. Нетрудно обнаружить аналогию со спектром гармонического колебания, балансно-модулированного детерминированным сигналом с полосовым спектром.

Если на основе балансно-модулированного сигнала образовать аналитический сигнал $\dot{s}(t)=[u(t)\cos\omega_0 t+iu(t)\sin\omega_0 t]/\sqrt{2}$, его спектральная плотность будет лежать в положительной области частот и будет равна

$$\Omega_0(\omega)=\frac{\alpha D}{\alpha^2+(\omega_0-\omega)^2},\quad \omega\geqslant 0. \tag{3.8}$$

С учетом (2.23) ширина спектра аналитического модулированного сигнала

$$\Delta\omega=\frac{1}{\Omega_m}\int\limits_{0}^{\infty}\Omega_0(\omega)\,d\omega=\frac{\pi D}{\Omega_0(\omega_0)}=\frac{\pi\alpha D}{D}=\alpha\pi. \tag{3.9}$$

Следовательно, спектр аналитического модулированного сигнала имеет такую же ширину, что и спектр полезного сигнала (см. (2.30)), но он расположен

в области ω_0. Использование (3.9) позволяет определить условие (2.59) узкополосности модулированного сигнала через параметры корреляционных функций полезного сигнала и переносчика:

$$\alpha\pi/\omega_0 = \alpha/2F_0 \ll 1, \qquad (3.10)$$

где F_0 — средняя линейная частота переносчика.

Если для балансно-модулированного сигнала условие (3.10) выполняется, он может рассматриваться как узкополосный сигнал со средней частотой F_0.

Таким образом, алгоритм определения корреляционной функции и спектральной плотности модулированных сигналов прост: для конкретных видов модуляции необходимо вычислить интеграл (3.5) и с помощью соотношения Хинчина — Винера определить спектральную плотность. Однако для многих видов модуляции практическая реализация этого алгоритма наталкивается на трудности вычислительного характера.

Контрольные вопросы

1. Как определить корреляционную функцию модулированного сигнала?
2. Запишите и поясните смысл выражения для корреляционной функции балансно-модулированного сигнала.
3. Запишите и поясните смысл выражения для спектральной плотности балансно-модулированного сигнала.
4. Как определяется условие узкополосности балансно-модулированного сигнала через параметры корреляционных функций полезного сигнала и переносчика?

3.3. СЛУЧАЙНЫЕ И ШУМОПОДОБНЫЕ СИГНАЛЫ-ПЕРЕНОСЧИКИ

Широкополосные (шумоподобные) и узкополосные случайные сигналы все чаще используют как сигналы-переносчики. Например, в оптических системах связи с некогерентным излучением сигналы-переносчики являются узкополосными гауссовскими процессами. В многоканальных и многоадресных системах применяют шумоподобные сигналы. Многоадресные системы обеспечивают связь между любой парой из множества источников и получателей сообщений, размещенных в пространстве независимо друг от друга. В отличие от многоканальной системы, в многоадресной системе сигналы различных абонентов поступают в общий тракт передачи сигналов без уплотнения — без образования из них группового сигнала.

Необходимым условием выделения требуемого полезного сигнала является взаимная ортогональность или слабая взаимная корреляция сигналов различных адресов. Это достигается для простых сигналов сдвигом во времени или по частоте, а для сложных (шумоподобных) — селекцией по форме. Сущность селекции сигналов по форме заключается в следующем. Определяют взаимную энергию (см. (2.11)) принятого сигнала $z(t)$ и передаваемых сигналов $s_i(t)$, $i=1, 2, \ldots$ (Для случайных переносчиков взаимная энергия пропорциональна взаимокорреляционной функции.) Принимают решение, что передавался тот сигнал, взаимная энергия которого с принятым максимальна. Например, если

$$\int_0^T s_i(t)\, z(t)\, dt > \int_0^T s_k(t)\, z(t)\, dt, \qquad (3.11)$$

принимается решение о том, что передавался сигнал $s_i(t)$. На практике используют слабокоррелированные сигналы, взаимная энергия которых намного меньше энергии каждого сигнала.

Применение шумоподобных сигналов позволяет ослабить влияние замираний в каналах с многолучевым распространением радиоволн. Например, если на вход приемника поступают сигналы $s(t)$ и $s(t-\tau)$ двух лучей со сдвигом на время τ, то мощность результирующего сигнала

$$P = \langle [s(t) + s(t-\tau)]^2 \rangle = 2\,[P_1 + K(\tau)], \qquad (3.12)$$

где P_1 — средняя мощность модулированного сигнала $s(t)$; $K(\tau)$ — функция корреляции $s(t)$. При выполнении условия $T/\tau \gg 1$ (T — время усреднения) с ростом τ функция корреляции стремится к нулю тем быстрее, чем шире спектр сигнала. Следовательно, при $T/\tau \gg 1$ и широкополосном сигнале

$$P \simeq 2P_1 = \mathrm{const} \tag{3.13}$$

и средняя мощность сигнала остается примерно постоянной, несмотря на замирания.

Шумоподобные сигналы относятся, как уже отмечалось, к сложным (многомерным), широкополосным сигналам, база которых $\nu = 2FT > 2$. Эти сигналы не являются случайными. Однако создаются они по таким алгоритмам, что их статистические характеристики близки к характеристикам белого шума: энергетический спектр почти равномерен, а корреляционная функция имеет узкий основной лепесток и небольшие побочные. Широкое применение получили дискретные шумоподобные сигналы, которые при большой базе позволяют обеспечить высокую точность синхронизации на передающем и приемном концах. Информационная посылка длительностью T делится на N бинарных сигналов длительностью $\tau_0 = T/N$ с полосой $F \simeq 1/\tau_0$ и базой $\nu > 2$. Последовательности длиной T являются кодовыми комбинациями, которые строят так, чтобы выполнялось условие селекции сигналов по форме (3.11).

Широкополосные радиосигналы образуют с помощью модуляции. Нормированная корреляционная функция огибающей этих сигналов имеет главный максимум в области $[-\tau_0, \tau_0]$ и боковые лепестки, амплитуды которых примерно равны $1/N$.

Эта функция напоминает корреляционную функцию реализации «шума» с полосой F, что и обусловило появление термина «шумоподобные сигналы». При увеличении N растет различимость шумоподобных сигналов, что позволяет обеспечивать высокую помехоустойчивость передачи информации даже в случаях, когда уровень помех выше уровня сигналов.

Наиболее распространенным является способ построения дискретных сигналов на основе линейных рекуррентных последовательностей. Среди этих последовательностей важную роль играют *псевдослучайные бинарные М-последовательности Хаффмена.* Они представляют совокупность N периодически повторяющихся символов s_i, каждый из которых может принимать одно из двух значений: $+1$ или -1. Значение s_i определяется взятым с противоположным знаком произведением значений двух или большего четного числа n предыдущих сигналов:

$$s_i = -\prod_{k=1}^{n} s_{i-k}, \quad i = \overline{n+1,\ N}; \quad n > k. \tag{3.14}$$

Если в роли исходной используется последовательность $\overline{s_1,\ s_n}$, то используя рекуррентное соотношение (3.14), можно построить неповторяющуюся элементарную последовательность $\{s_i\}$ из N символов, где

$$N = 2^n - 1. \tag{3.15}$$

Эта последовательность будет содержать все комбинации n символов из двух элементов $(+1, -1)$, кроме комбинаций, которые состоят из одних отрицательных единиц. Поэтому каждая последовательность включает 2^{n-1} положительных единиц и $2^{n-1}-1$ отрицательных единиц и сумма

$$\sum_{i=1}^{N} s_i = 1. \tag{3.16}$$

Число N является максимальным периодом бесконечной последовательности Хаффмена. Можно образовать последовательности и меньшего периода. Если число символов в сигнале больше N, символы повторяются в соответствии с рекуррентным соотношением (3.14), порождающим эту последовательность. Мак-

симальное число $M_п$ различных последовательностей максимального периода для любого n равно

$$M_п = \frac{1}{n} \varphi (2^n - 1), \qquad (3.17)$$

где $\varphi(\cdot)$ — функция Эйлера, равная количеству целых чисел, включая единицу, меньших 2^n-1 и взаимно-простых с числом 2^n-1.

Двоичные псевдослучайные последовательности Хаффмена обладают рядом практически полезных свойств: возможностью формирования обширного множества квазиортогональных сигналов, простотой построения сигналов и хорошими корреляционными свойствами сигналов. При непрерывном излучении нормированная корреляционная функция огибающей шумоподобного сигнала, построенного на основе таких последовательностей, имеет главный максимум, равный единице, в области $[-\tau_0, \tau_0]$, где τ_0 — длительность сигналов s_i, и одинаковые по величине боковые лепестки, амплитуда которых равна $-1/N$.

Взаимокорреляционная функция для различных последовательностей равна $-1/M$. При выбранном n различные последовательности отличаются как порядком чередования символов $+1$ и -1, так и максимальным значением амплитуд боковых лепестков корреляционной функции корреляции. Обычно отыскивают такие последовательности, для которых максимальный уровень боковых лепестков будет наименьшим среди всех последовательностей, порождаемых числом n. Генерирование псевдослучайных последовательностей напоминает процедуру генерирования случайных чисел в ЦВМ и легко осуществляется с помощью регистров сдвига (см. § 8.2). Наряду с сигналами Хаффмена применяют и другие шумоподобные сигналы.

Особенности образования последовательностей Хаффмена рассмотрим на примере. Выберем $n=2$, тогда из (3.14) получим

$$s_i = -s_{i-2}s_{i-1}. \qquad (3.18)$$

Для использования рекуррентного соотношения (3.18) выберем исходную последовательность $s_1s_2=-1+1$. Из (3.18), находим

$$s_3=+1,\ s_4=-1,\ s_5=+1,\ s_6=+1, \ldots$$

Искомая последовательность имеет вид: $-1+1+1-1+1+1-1+1+1 \ldots$ Период этой последовательности $N=2^2-1=3$, поэтому набор символов $-1+1+1$ повторяется через три символа. Так как $M=1$, то при $n=2$ может быть образована только одна последовательность. Если выбрать $n=3$, можно образовать две последовательности ($M=2$) с периодом $N=7$ по правилам

$$s_i=-s_{i-3}s_{i-2},\quad s_i=-s_{i-3}s_{i-1}.$$

Шумоподобные сигналы можно подвергать всем простейшим видам параметрической модуляции. При амплитудной модуляции изменяются амплитуды всех элементов шумоподобного сигнала, при частотной — сигналы отличаются средней частотой, при фазовой — разностью фаз между элементами двух посылок. Однако наибольшего внимания заслуживает модуляции шумоподобных сигналов по форме (структурная модуляция). При этой модуляции в качестве сигналов используют различные последовательности, образованные с помощью соотношения (3.14).

Сигналы содержат одинаковые элементы $(+1, -1)$, но расположение их внутри составного сигнала определяется соотношением (3.14). Например, двоичную передачу можно осуществить с помощью двух широкополосных сигналов $s_1(t)=u_2(t)\cos\omega_0 t$, $s_2(t)=u_2(t)\cos\omega_0 t$, где в роли огибающих $u_1(t)$ и $u_2(t)$ используются последовательности Хаффмена $u(t)$, сдвинутые на время τ_0, $u_1(t)= =u(t)$, $u_2(t)=u(t+\tau_0)$. Тогда

$$\int_0^T u_1(t)\, u_2(t)\, dt = \int_0^T u(t)\, u(t+\tau_0)\, dt \leqslant \frac{1}{N}.$$

Аналогично строят многопозиционные системы, в которых используют m квазиортогональных шумоподобных сигналов.

Контрольные вопросы

1. Где применяются широкополосные и узкополосные случайные сигналы-переносчики?
2. В чем сущность селекции сигналов по форме?
3. Почему шумоподобные сигналы позволяют ослабить влияние замирания сигналов?
4. Какими корреляционными и спектральными характеристиками обладают шумоподобные сигналы?
5. Укажите особенности линейных рекуррентных последовательностей.
6. Укажите полезные свойства дискретных сигналов, построенных на основе двоичных псевдослучайных последовательностей Хаффмена.
7. Какие виды модуляции шумоподобных сигналов используют?
8. Как осуществить двоичную передачу с помощью широкополосных сигналов?

3.4. АМПЛИТУДНАЯ МОДУЛЯЦИЯ СЛУЧАЙНОГО СИГНАЛА

Особенности непрерывной амплитудной модуляции (АМ) (модуляция класса С3) (см. § 3.1) случайного процесса рассмотрим при следующих предположениях. Модулирующий сигнал $u(t)$ является случайным стационарным процессом с математическим ожиданием $M[u(t)]=0$ и дисперсией $M[u^2(t)]=\sigma^2$, корреляционная функция этого сигнала $K_1(\tau)$. Случайный переносчик $x(t)$ также является стационарным процессом с корреляционной функцией $K_2(\tau)$. Процессы $u(t)$ и $x(t)$ — независимы. Определим корреляционную функцию и спектральную плотность амплитудно-модулированного случайного сигнала.

Амплитудно-модулированный сигнал $s(t)=x(t)[1+mu(t)]$, где m — коэффициент модуляции.

Корреляционную функцию модулированного сигнала получим, применяя операции усреднения по времени и по множеству. Тогда

$$K(\tau)=M[\langle s(t)\, s(t+\tau)\rangle]=M\{\langle[1+1+mu(t)][1+mu(t+\tau)]\, x(t)\, x(t+\tau)\rangle\}=$$

$$=M\{\langle[1+mu(t)+mu(t+\tau)+m^2u(t)\, u(t+\tau)]\, x(t)\, x(t+\tau)\rangle\}.$$

После выполнения операций усреднения получим

$$K(\tau)=K_2(\tau)+m^2K_1(\tau)K_2(\tau)=K_2(\tau)[1+m^2K_1(\tau)]. \qquad (3.19)$$

Сравнение (3.6) и (3.19) показывает, что корреляционная функция амплитудно-модулированного случайного сигнала представляет сумму корреляционной функции переносчика и корреляционной функции балансно-модулированного случайного сигнала. Следовательно, зная характеристики балансно-модулированного случайного сигнала, можно определить корреляционную функцию амплитудно-модулированного случайного сигнала.

Используя соотношение Хинчина — Винера, найдем спектральную плотность амплитудно-модулированного случайного сигнала

$$\Omega(\omega)=2\int_0^\infty K_2(\tau)\cos\omega\tau d\tau+2\int_0^\infty m^2K_1(\tau)\, K_2(\tau)\cos\omega\tau d\tau. \qquad (3.20)$$

Первый интеграл в (3.20) определяет спектральную плотность случайного переносчика, а второй — спектральную плотность балансно-модулированного сигнала. Используя теорему о спектре произведения корреляционных функций, формулу (3.20) можно представить в виде

$$\Omega(\omega)=\Omega_2(\omega)+m^2\int_{-\infty}^{\infty}\Omega_1(\omega-\nu)\,\Omega_2(\nu)\,d\nu, \qquad (3.21)$$

где через Ω_1 обозначена спектральная плотность полезного сигнала.

Найдем спектральную плотность амплитудно-модулированного сигнала для случая, когда

$$K_1(\tau) = 0{,}5\mathcal{D}_1 e^{-\alpha_1|\tau|}\cos\omega_1\tau,$$

$$K_2(\tau) = 0{,}5\mathcal{D}_2 e^{-\alpha_2|\tau|}\cos\omega_2\tau, \quad \omega_2 > \omega_1.$$

Второй интеграл в (3.20) можно привести к виду

$$m^2\int_{-\infty}^{\infty} K_1(\tau)K_2(\tau)e^{-i\omega\tau}d\tau = \frac{m^2}{4}\int_{-\infty}^{\infty}\mathcal{D}_1\mathcal{D}_2 e^{-(\alpha_1+\alpha_2)\tau}\cos\omega_1\tau\cos\omega_2\tau e^{-i\omega\tau}d\tau.$$

Воспользуемся формулой для произведения косинусов и представим этот интеграл в виде суммы двух интегралов

$$\frac{m^2\mathcal{D}_1\mathcal{D}_2}{8}\left[\int_{-\infty}^{\infty} e^{-\alpha\tau}\cos(\omega_2+\omega_1)\tau e^{-i\omega\tau}d\tau + \int_{-\infty}^{\infty} e^{-\alpha\tau}\cos(\omega_2-\omega_1)\tau e^{-i\omega\tau}d\tau\right],$$

где $\alpha=\alpha_1+\alpha_2$.

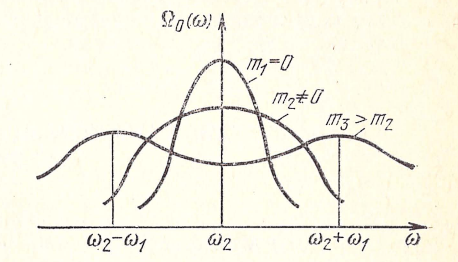


Рис. 3.1. Графики зависимости $\Omega_0(\omega)$ при различных значениях m

Такие интегралы ранее вычислялись в § 3.2 при определении спектра балансно-модулированного гармонического колебания. Использовав эти результаты, после необходимых преобразований окончательно получим

$$\Omega(\omega) = \mathcal{D}_2\alpha_2\left[\frac{1}{\alpha_2^2+(\omega-\omega_2)^2}+\frac{1}{\alpha_2^2+(\omega+\omega_2)^2}\right]+$$

$$+\frac{m^2\mathcal{D}_1\mathcal{D}_2\alpha}{4}\left[\frac{1}{\alpha^2+(\omega-\omega_2+\omega_1)^2}+\frac{1}{\alpha^2+(\omega+\omega_2-\omega_1)^2}\right]+$$

$$+\frac{m^2\mathcal{D}_1\mathcal{D}_2\alpha}{4}\left[\frac{1}{\alpha^2+(\omega-\omega_2-\omega_1)^2}+\frac{1}{\alpha^2+(\omega+\omega_2+\omega_1)^2}\right].$$

Аналитический модулированный сигнал, соответствующий рассматриваемому действительному амплитудно-модулированному сигналу, имеет спектральную плотность

$$\Omega_0(\omega) = \frac{2\alpha_2\mathcal{D}_2}{\alpha_2^2+(\omega_2-\omega)^2}+\frac{m^2\mathcal{D}_1\mathcal{D}_2\alpha}{2[\alpha^2+(\omega_2+\omega_1-\omega)^2]}+\frac{m^2\mathcal{D}_1\mathcal{D}_2\alpha}{2[\alpha^2+(\omega_2-\omega_1-\omega)^2]}. \tag{3.22}$$

На рис. 3.1 показаны графики $\Omega_0(\omega)$ при различных значениях коэффициента модуляции m. Можно заметить аналогию между спектром случайного переносчика, амплитудно-модулированного случайным сигналом, и спектром амплитудно-модулированного детерминированного колебания [3]. И в том и в другом случае есть несущая и две боковые полосы частот. Отличие заклю-

чается в том, что спектр амплитудно-модулированного случайного сигнала является непрерывным, роль несущей играет узкополосный случайный переносчик, боковые полосы отстоят от средней частоты ω_2 этого сигнала на величину ω_1 (ω_1 — средняя частота модулирующего сигнала). Ширина спектра боковых полос определяется шириной спектра переносчика и шириной спектра полезного сигнала.

Следовательно, зная корреляционные функции переносчика и полезного сигнала, а также оператор модуляции, можно определить корреляционную функцию и спектральную плотность случайного амплитудно-модулированного сигнала. Корреляционные и спектральные характеристики модулированных случайных сигналов при частотной и фазовой модуляциях определяют аналогично. Для упрощения аналитических преобразований случайный переносчик обычно представляют в виде узкополосного процесса.

Контрольные вопросы

1. Запишите и поясните аналитическое выражение для амплитудно-модулированного случайного сигнала.
2. Запишите и поясните аналитическое выражение для корреляционной функции амплитудно-модулированного случайного сигнала.
3. Нарисуйте и поясните графики спектральной плотности аналитического амплитудно-модулированного сигнала при различных значениях коэффициента модуляции.

3.5. АМПЛИТУДНАЯ МАНИПУЛЯЦИЯ И АМПЛИТУДНАЯ ИМПУЛЬСНАЯ МОДУЛЯЦИЯ

Рассмотрение амплитудной манипуляции (АМн) и амплитудной импульсной модуляции (АИМ) позволяет проиллюстрировать особенности анализа модулированных сигналов при дискретной и импульсной модуляциях.

3.5.1. Амплитудная манипуляция (АМн). В этом виде дискретной модуляции в роли информационного параметра переносчика выступает амплитуда, которая изменяется скачкообразно под воздействием модулирующего сигнала (отсюда и название — амплитудная манипуляция) [3]. Амплитудная манипуляция относится к классу **В1**.

Рассмотрим особенности анализа АМн сигнала для случая, когда в роли переносчика выступает гармоническое колебание $x(t)=A_0 \sin(\omega_2 t+\varphi_0)$, а в роли модулирующего сигнала — периодическая последовательность прямоугольных импульсов

$$u(t)=\begin{cases} 1, & 2i\tau<t<(2i+1)\tau, \\ -1, & (2i+1)\tau<t<2(i+1)\tau, \quad i=0,\ 1,\ 2, \ldots, \end{cases}$$

где τ — длительность импульсов; $T=2\tau$ — период последовательности. В этом случае амплитуда манипулированного сигнала принимает два значения:

$$A=\begin{cases} 0{,}5A_0(1+m), & 2i\tau<t<(2i+1)\tau, \\ 0{,}5A_0(1-m), & (2i+1)\tau<t<2(i+1)\tau, \quad i=0,\ 1,\ 2, \ldots \end{cases}$$

Обычно коэффициент модуляции m при АМн выбирают равным единице. Поэтому амплитуда манипулированного сигнала изменяется скачком в моменты времени $t=i\tau$, $i=0,\ 1,\ 2, \ldots$, и принимает два значения: A_0 и 0. На рис. 3.2 показаны временные диаграммы модулирующего $u(t)$ и манипулированного $s(t)$ сигналов. Можно заметить, что при АМн источник высокочастотных колебаний работает в режиме прерывистой генерации.

Аналитически АМн сигнал записывают так:

$$s(t)=0{,}5A_0[1+u(t)]\sin(\omega_2 t+\varphi_0). \tag{3.23}$$

Определим спектр этого сигнала. Представим $u(t)$ в виде ряда Фурье

$$u(t) = \frac{2}{\pi} \sum_{k=1}^{\infty} \frac{1 - \cos k\pi}{k} \sin k\omega_1 t, \qquad (3.24)$$

где $\omega_1 = 2\pi/T$. Подставив (3.24) в (3.23), получим

$$s(t) = A_0 \left(\frac{1}{2} + \frac{1}{\pi} \sum_{k=1}^{\infty} \frac{1 - \cos k\pi}{k} \sin k\omega_1 t \right) \sin(\omega_2 t + \varphi_0) =$$

$$= \frac{1}{2} A_0 \sin(\omega_2 t + \varphi_0) + \frac{A_0}{2\pi} \sum_{k=1}^{\infty} \frac{1 - \cos k\pi}{k} \cos(\omega_2 t - k\omega_1 t + \varphi_0) -$$

$$- \frac{A_0}{2\pi} \sum_{k=1}^{\infty} \frac{1 - \cos k\pi}{k} \cos(\omega_2 t + k\omega_1 t + \varphi_0). \qquad (3.25)$$

На рис. 3.3 показан построенный по формуле (3.25) спектр АМн сигнала. Огибающая спектра (штриховая линия) представляет смещенный на частоту ω_2 спектр одиночного видеоимпульса $u(t)$. С очевидными изменениями полученные результаты справедливы для более общих случаев, когда $u(t)$ является

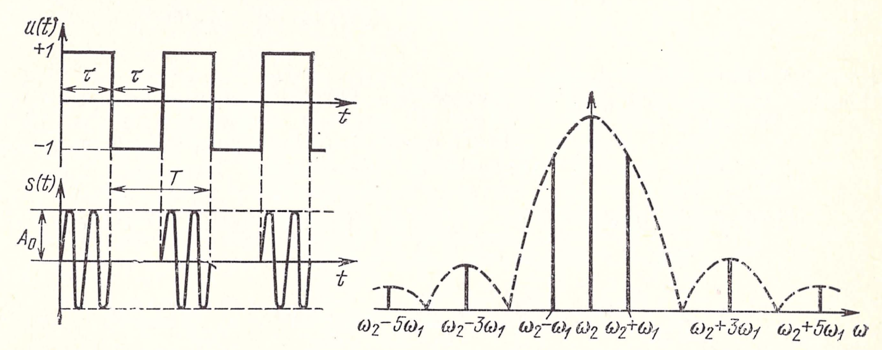


Рис. 3.2. Временные диаграммы модулирующего и манипулированного сигналов

Рис. 3.3 Спектр АМн сигнала

случайной последовательностью знакопеременных импульсов с детерминированной амплитудой (см. § 2.4), когда $x(t)$ является случайным сигналом и когда $u(t)$ и $x(t)$ являются случайными. Для анализа АМн сигналов в таких более общих случаях применяют алгоритмы, приведенные в § 3.2—3.4.

Например, спектр гармонического колебания, амплитудно-манипулированного случайным телеграфным сигналом с корреляционной функцией (2.27) (модуляция класса D1), имеет вид

$$\Omega(\omega) = \frac{A_0^2}{4} \left[\delta(\omega_2 - \omega) + \frac{\sigma^2 \alpha}{\alpha^2 + (\omega_2 - \omega)^2} \right], \qquad (3.26)$$

где $\delta(\omega_2 - \omega)$ — дельта-функция [3, 15]. Следовательно, при АМн спектр (2.29) случайного телеграфного сигнала переносится на частоту ω_2 и «накладывается» на спектральную линию гармонического колебания. Ширина спектра по-прежнему определяется соотношением (2.30).

3.5.2. **Амплитудная импульсная модуляция (АИМ).** При АИМ роль переносчика выполняет периодическая последовательность видеоимпульсов:

$$x(t) = A_0 \sum_{j=-\infty}^{\infty} x_1(t - jT,\ \tau),$$

где A_0 — амплитуда импульсов; $x_1(t)$ — функция, описывающая одиночный импульс последовательности; T — период повторения импульсов; τ — длительность одного импульса. На рис. 3.4 показаны временные диаграммы модулирующего сигнала $u(t)$ и АИМ сигнала. Штриховой линией обозначена немодулированная импульсная видеопоследовательность.

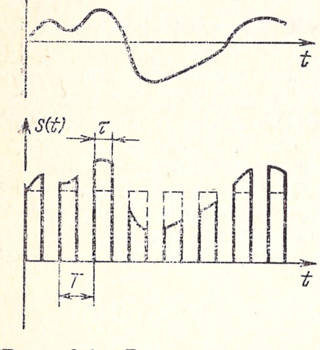


Рис. 3.4. Временные диаграммы модулирующего и АИМ сигнала

Аналитическая запись АИМ сигнала (класс А2) имеет вид

$$s(t) = A_0[1 + mu(t)] \sum_{j=-\infty}^{\infty} x_1(t - jT,\ \tau), \tag{3.27}$$

где m — коэффициент модуляции. Определим спектр $s(t)$. Представим $x(t)$ в виде ряда Фурье

$$x(t) = \sum_{k=-\infty}^{\infty} \dot{c}_k e^{ik\omega_2 t}, \tag{3.28}$$

где $\omega_2 = 2\pi/T$ — круговая частота повторения импульсов. Подставляя значение $x(t)$ из (3.28) в (3.27) и используя преобразование Фурье, находим спектр АИМ сигнала

$$\Omega(i\omega) = \int_{-\infty}^{\infty} s(t) e^{-i\omega t} dt = \int_{-\infty}^{\infty} [1 + mu(t)] \sum_{k=-\infty}^{\infty} \dot{c}_k e^{-i(\omega - k\omega_2)t} dt =$$

$$= 2\pi \sum_{k=-\infty}^{\infty} \dot{c}\delta(\omega - k\omega_2) + m \sum_{k=-\infty}^{\infty} \dot{c}_k \int_{-\infty}^{\infty} u(t\ e^{-i(\omega - k\omega_2)t} dt. \tag{3.29}$$

Первая сумма в формуле (3.29) представляет спектр немодулированной последовательности (3.28). Вторая сумма показывает, что амплитудная модуляция вызывает появление возле каждой составляющей этого спектра боковых полос, повторяющих спектр модулирующего сигнала. Поэтому спектр АИМ сигнала представляет упорядоченный набор спектров обычных АМ колебаний, в которых роль несущих выполняют гармоники частоты следования видеоимпульсов. Для иллюстрации особенностей АИМ на рис. 3.5 показан типичный вид спектра АИМ сигнала для случая, когда $u(t)$ является узкополосным случайным сигналом (см. § 2.6) со средней частотой ω_1. Штриховой линией показана огибающая спектра немодулированной последовательности видеоимпульсов. Следует уточнить, что в данном случае спектр АИМ сигнала определен не с помощью преобразования Фурье (3.29), а с помощью преобразования Хинчина — Винера (3.3), так как рассматривается модуляция класса С2.

Рассмотрение спектра АИМ сигнала позволяет сделать ряд практически важных выводов. Очевидно, что необходимо выбирать такую минимальную частоту повторения импульсов

$$\omega_{2\text{мин}} \geqslant 2\omega_1, \tag{3.30}$$

при которой не происходит наложения спектров соседних боковых полос. Если условие (3.30) выполняется, можно выделить составляющие модулированного

сигнала с помощью полосовых фильтров и фильтров нижних частот. Практически важной особенностью спектра АИМ сигнала (она проявляется и в других видах импульсной модуляции) является наличие около частоты $\omega=0$ составляющих модулирующего сигнала (рис. 3.5). Следовательно, демодуляцию АИМ сигнала можно выполнить фильтром нижних частот без дополнительных преобразований. Фильтр должен пропускать частоты от 0 до $\omega_0 \approx \omega_2 - \omega_1$ (рис. 3.5).

Частоте $\omega_{2мин}$ соответствует период $T_{макс}$. Большие временны́е интервалы между импульсами используют для размещения импульсов других каналов при многоканальной передаче с временны́м разделением сигналов (см.

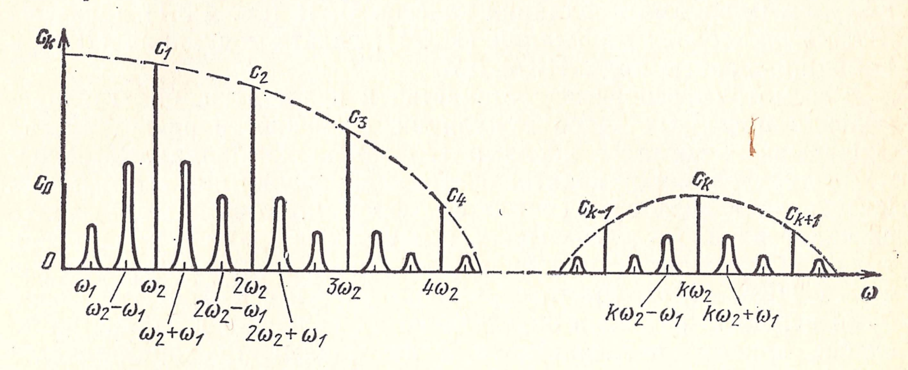


Рис. 3.5. Спектр АИМ сигнала

§ 9.2). Длительность τ импульсов определяется полосой пропускания каналов. Величину T/τ называют скважностью, обычно $T/\tau \gg 1$. Чаще всего АИМ видеосигнал используют как модулирующий сигнал для создания высокочастотных модулированных колебаний. На первом этапе образуют АИМ сигнал, а на втором — полученный АИМ видеосигнал применяют для модуляции непрерывного высокочастотного переносчика, имеющего частоту $\omega_3 \gg \omega_2$. После таких преобразований спектр сигнала $u(t)$ переносится на частоту ω_3 несущего высокочастотного колебания. Анализ модулированных высокочастотных колебаний выполняют с учетом вида модуляции методами, изложенными в § 3.2—3.5.

Контрольные вопросы

1. В чем сущность амплитудной манипуляции и амплитудной импульсной модуляции?
2. Поясните временные диаграммы и спектры сигналов при амплитудной, частотной и фазовой манипуляции.
3. Поясните временные диаграммы и спектры сигналов при амплитудной, широтной и фазовой импульсной модуляции.

3.6. ЦИФРОВЫЕ МЕТОДЫ МОДУЛЯЦИИ

Цифровые виды модуляции используют для передачи непрерывных сообщений дискретными методами. Сущность цифровой модуляции заключается в том, что передаваемый непрерывный сигнал дискретизируется во времени, квантуется по уровню и полученные после этих операций отсчеты, следующие в дискретные моменты времени, рассматриваются в той или иной системе счисления как числа, которые затем кодируются для преобразования их в кодовые комбинации электрических сигналов. Полученной последова-

тельностью кодовых видеосигналов аналоговым или дискретным способом модулируют высокочастотный сигнал-переносчик.

Следовательно, цифровые методы модуляции основаны на трех необходимых преобразованиях полезных непрерывных сигналов: дискретизации, квантовании и кодировании. Четвертое преобразование — модуляцию — используют, как правило, при передаче сигналов в многоканальных системах. Наиболее распространенными цифровыми видами модуляции классов **G**, **H** (см. § 3.1) являются *импульсно-кодовая модуляция* (ИКМ), дельта-модуляция (ДМ) и комбинированные виды ИКМ-ДМ.

Рассмотрим основные достоинства и недостатки, области применения цифровых видов модуляции, особенности представления непрерывных сигналов числами (кодовыми комбинациями), сущность дискретизации, квантования и кодирования при ИКМ и ДМ. Интенсивное внедрение систем с цифровыми способами передачи сигналов началось в 60-х годах, в настоящее время в мире насчитывается несколько миллионов эффективно эксплуатируемых систем с ИКМ [5, 8]. Быстрое развитие систем с цифровыми способами передачи сигналов обусловлено их важными достоинствами: слабое влияние неидеальности и нестабильности характеристик аппаратуры на качество передачи информации; высокая помехоустойчивость даже при использовании каналов с нестабильными характеристиками и большим уровнем шумов; возможность регенерации (восстановления) сигналов в узлах связи сетей, что значительно ослабляет эффект накопления искажений сигналов при передаче информации по линиям большой протяженности; универсальная форма представления сигналов для различных сообщений (речь, телевизионные изображения, дискретные данные, команды управления работой устройства связи и т. п.); низкая чувствительность к нелинейным искажениям в групповом тракте многоканальных систем; относительно простое согласование этих систем с ЦВМ и с электронными автоматическими телефонными станциями, что играет важную роль для построения интегральных сетей связи; возможность автоматизации передачи и обработки сигналов с помощью ЦВМ; возможность унификации и стандартизации элементов и устройств систем и сетей связи на основе интегральной микросхемотехники и цифровой вычислительной техники; применение малогабаритных цифровых фильтров для селекции сигналов.

Основными недостатками систем с цифровыми способами передачи сигналов являются: значительное расширение занимаемой полосы частот каналов, необходимость обеспечения точной синхронизации сигналов и построения аппаратуры для их регенерации на линиях большой протяженности. Например, при одинаковом числе каналов для передачи группового сигнала с ИКМ и временным разделением сигналов для многоканальной системы требуется полоса частот примерно в 15 раз большая, чем в системе с частотным разделением сигналов и однополосной модуляцией [8]. Фазовые ошибки тактовой синхронизации ограничивают максималь-

ную длину тракта передачи информации. Результаты исследований последних лет показывают, что эти недостатки не играют определяющей роли и цифровые методы передачи информации будут применяться все более широко.

Наметились следующие основные области применения цифровых видов модуляции: уплотнение действующих соединительных линий между АТС, построение интегральных сетей связи с электронными АТС, радиорелейные линии связи прямой видимости и дальние линии связи, спутниковые системы связи с многостанционным доступом, волноводные и оптические линии многоканальной связи, телеметрия. В настоящее время наибольшее развитие получили системы с ИКМ для уплотнения телефонных линий связи и в спутниковой связи. Принципиальная возможность применения цифровых видов модуляции обусловлена следующими причинами: абсолютно точная передача непрерывных сигналов невозможна из-за искажений и шумов в каналах связи, поэтому *погрешности дискретизации и квантования сигналов* являются допустимыми; разрешающая способность и точность работы устройств и операторов ограничены; непрерывные сигналы обладают значительной избыточностью.

Последовательно рассмотрим сущность операций дискретизации, квантования и кодирования применительно к цифровой передаче непрерывных сигналов.

На рис. 3.6 показаны временные диаграммы сигналов в системе с ИКМ. На рис. 3.6,*а* показаны исходный непрерывный сигнал с ограниченным спек-

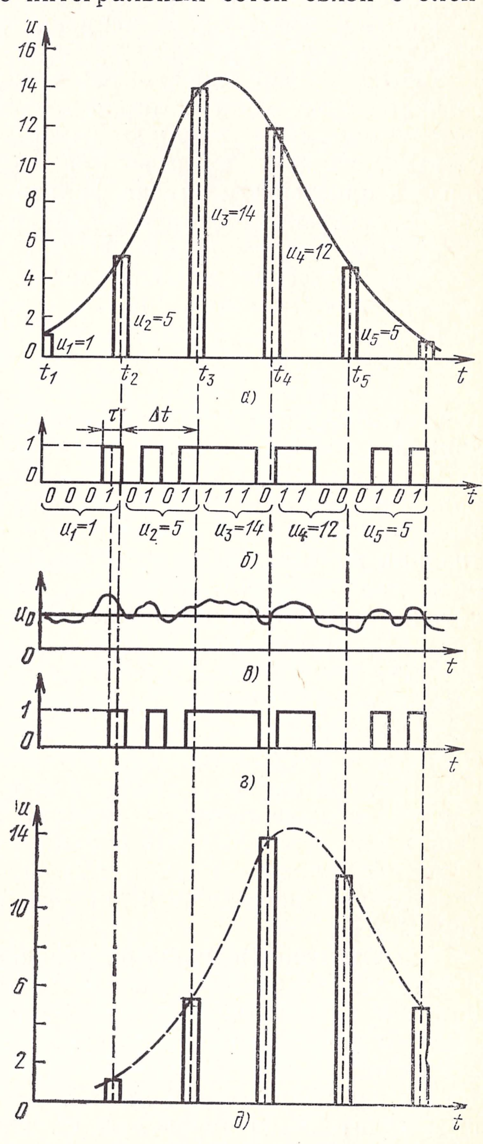


Рис. 3.6. Временные диаграммы сигналов в системе ИКМ

тром и его дискретизация с интервалом $\Delta t=1/2F$, где F — верхняя частота спектра сигнала, на рис. 3.6,*б* — полученная в результате квантования и кодирования последовательность двоичных видеосигналов. Из-за искажений сигналов и шумов в канале принятая видеопоследовательность (рис. 3.6,*в*) отличается от переданной. Выбирается порог u_0. Его превышение в моменты отсчета (стробирования) значениями сигнала означает наличие импульса, а непревышение — отсутствие; с помощью формирующих устройств из принятой видеопоследовательности создается «очищенная» — регенерированная последовательность сигналов (рис. 3.6,*г*), которая поступает на декодер. С выхода декодера импульсы, площадь которых равна соответствующим отсчетам исходного сигнала (рис. 3.6,*д*), поступают на демодулятор — в простейшем случае фильтр нижних частот, на выходе которого восстанавливается копия исходного непрерывного сигнала (штриховая линия на рис. 3.6,*д*).

Для получения регенерированной кодовой последовательности отсчеты принимаемого сигнала берут в середине каждого тактового интервала τ (рис. 3.6,*б* и *в*). Это делается для того, чтобы исключить влияние на работу демодулятора запаздывания и фазовых искажений сигналов в канале связи. В результате регенерируемая последовательность «задержана» на $\tau/2$ относительно переданной (рис. 3.6,*б* и *г*). Правильное декодирование сигналов требует также, чтобы были приняты все разряды кодовой комбинации. Из-за этого принятые отсчеты оказываются дополнительно задержанными относительно передаваемых на интервал дискретизации Δt (рис. 3.6,*а* и *д*). Важную роль играет идентификация разрядов кодовых комбинаций, выполняемая счетчиками импульсов, которые делят последовательность видеоимпульсов на комбинации требуемой длины. Для синхронизации работы счетчиков периодически передают легко индентифицируемую (тестовую, контрольную) кодовую комбинацию.

Дискретизация во времени — это преобразование непрерывного сигнала в дискретную последовательность отсчетов (рис. 3.6,*а*). Она осуществляется выбором отсчетов $u_k=u(t_k)$ сигнала $u(t)$ в определенные моменты времени $t_k=k\Delta t$, $k=1, 2, \ldots$ Интервал (шаг) дискретизации, как правило, выбирают в соответствии с теоремой Котельникова (см. § 2.2). Погрешности дискретизации определяют как погрешности ортогональных разложений сигналов (см. § 2.1).

Квантование сигналов по уровню — это дискретизация полученной последовательности отсчетов по уровню. Сущность квантования как нелинейного преобразования заключается в том, что все отсчеты, попадающие в интервал квантования Δu_k, представляют одним значением u^*_k, которое называют квантованным (рис. 3.7). Квантованные значения сигнала указаны на рис. 3.7 точками. Процесс квантования определен, если задана *характеристика квантования* (рис. 3.8). Она связывает интервалы квантования и квантованные значения: каждому интервалу квантования Δu_k ставится

в соответствие квантованное значение u^*_k. Часто интервалы квантования выбирают одинаковыми и говорят, что квантование производится с постоянным шагом. Если выбирать $u^*_k=(u_k+u_{k+1})/2$, можно упростить схему устройства квантования (квантователя).

Характерной особенностью цифровых видов модуляции является то, что из-за погрешностей квантования копия сигнала отличается от оригинала даже при полном отсутствии помех и искажений в канале. Поэтому вводят такую характеристику, как *шумы квантования* $\varepsilon(t)$ и для копии сигнала используют выражение

$$z(t)=u(t)+\varepsilon(t). \tag{3.31}$$

Если квантование выполняют с равномерным шагом, то область изменения мгновенных значений шума квантования (абсолютной погрешности квантования) $\varepsilon(t)$ заключена в пределах от $-\Delta u/2$ до $\Delta u/2$. На рис. 3.7 показано значение абсолютной погрешности

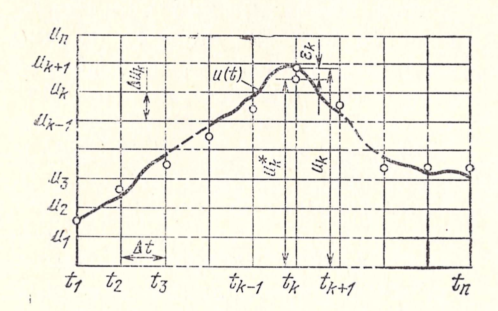


Рис. 3.7. Квантование уровня отсчетов

Рис. 3.8. Характеристика квантования

квантования $\varepsilon_k=u_k-u^*_k$ для момента дискретизации t_k. Обычно ошибку квантования считают равномерно распределенной в интервале $[-\Delta u/2, \Delta u/2]$, поэтому математическое ожидание и дисперсия ошибки

$$M[\varepsilon(t)]=0, \quad D[\varepsilon(t)]=\int\limits_{-\Delta u/2}^{\Delta u/2} \varepsilon^2 \frac{d\varepsilon}{\Delta u}=\frac{\Delta u^2}{12}. \tag{3.32}$$

Дисперсию ошибки квантования рассматривают как мощность шумов квантования (мощность помех, обусловленных квантованием). Мощность шумов квантования падает с ростом числа уровней квантования и при правильном выборе этого числа может быть пренебрежимо мала по сравнению с мощностью помех в канале. Обычно при квантовании решают две основные задачи: анализ погрешностей квантования (анализ искажений полезного сигнала при квантовании) и оптимизация квантования с учетом помех в канале (оптимальный синтез квантователя, обеспечивающего минимальные погрешности). На практике во многих системах, например в многоканальных системах с ИКМ, равномерное квантование не применяют. Это обусловлено рядом причин: рас-

пределение ошибки квантования отличается от равномерного; динамический диапазон передаваемых сигналов различен; для обеспечения максимальной скорости передачи информации стремятся обеспечить появление квантованных значений с одинаковой вероятностью (см. § 5.1).

Погрешности квантования можно уменьшить и при равномерном квантовании. Для этого необходимо сжимать динамический диапазон полезных сигналов. Такое сжатие называют компрессированием динамического диапазона. При этом уменьшается интервал квантования Δu и соответственно падает мощность шумов квантования (3.32). Компрессирование сигналов выполняют перед квантованием. После декодирования и восстановления сигнала необходимо выполнить обратную операцию, называемую экспандированием. Обе операции называют компандированием сигнала. Таким образом, влияние шумов квантования и помех можно уменьшить, оптимально выбирая параметры процесса квантования и искусственно изменяя динамический диапазон полезных сигналов (применяя компандирование). Управляемыми переменными при квантовании являются уровни квантования, интервалы квантования и квантованные значения.

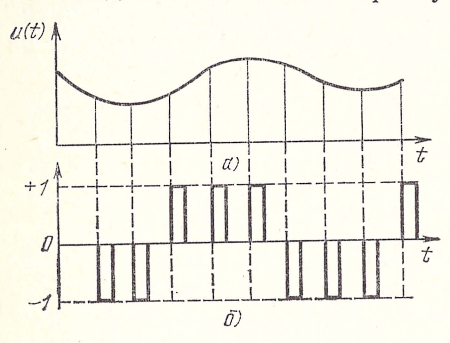


Рис. 3.9. Временные диаграммы при дельта-модуляции

Более простой по сравнению с ИКМ является дельта-модуляция (ДМ). Она позволяет применить элементарный способ декодирования сигналов и использовать меньшую полосу частот. Сущность ДМ заключается в том, что в дискретный момент времени взятия отсчета передается положительный импульс, если производная сигнала в этой точке положительна, и отрицательный импульс, если она отрицательна (рис. 3.9). В результате кодирования в дельта-модуляторе непрерывный сигнал преобразуется в последовательность положительных и отрицательных видеоимпульсов (рис. 3.9,*б*). Так же, как и при ИКМ, эта последовательность в случае необходимости может быть использована для модуляции высокочастотного сигнала-переносчика. Декодер выполняют как простой интегрирующий контур, который суммирует все импульсы кодовой видеопоследовательности. На выходе контура получают копию полезного сигнала. Системы с ДМ обладают многими достоинствами систем с ИКМ, но применяемая в них аппаратура кодирования и декодирования является более простой. Поэтому разрабатывают и применяют комбинированные виды модуляции: ИКМ-ДМ.

Для кодирования N квантованных значений сигнала при ИКМ с помощью примитивного (некорректирующего) кода необходимо так выбирать основание кода m и значность кода n, чтобы $m^n = N$.

Если обеспечена одинаковая вероятность появления квантованных значений и интервал Δt дискретизации выбран так, что они независимы, максимальная скорость передачи информации при ИКМ

$$R_{\text{макс}} = (n \log m)/\Delta t = 2Fn \log m, \qquad (3.33)$$

где $2F$ — частота дискретизации полезного сигнала. Для двоичного кодирования максимальная скорость

$$R_{2\text{макс}} = 2Fn. \qquad (3.34)$$

В современных системах с цифровыми способами передачи сигналов начинают применять корректирующее кодирование. Это приводит к некоторому снижению скорости передачи информации, но увеличивает помехоустойчивость.

Контрольные вопросы

1. В чем сущность цифровых способов передачи непрерывных сигналов?
2. Назовите достоинства и недостатки систем с ИКМ и ДМ.
3. Где применяют цифровые виды модуляции?
4. На каких операциях основано применение цифровых методов модуляции?
5. В чем сущность ИКМ? Нарисуйте и поясните временные диаграммы.
6. В чем сущность ДМ?
7. Какие особенности использования ИКМ?
8. Что такое шумы квантования и как оценивают их мощность?
9. Почему применяют неравномерное квантование и компандирование сигналов?
10. Как определяется скорость передачи информации в системах с ИКМ?

3.7. ВЫВОДЫ

1. Анализ видов модуляции и характеристик модулированных сигналов выполняют для согласования характеристик сигналов и каналов, сокращения естественной избыточности сигналов, определения потенциальной помехоустойчивости различных видов модуляции, определения помех соседним каналам.

Модулированные сигналы являются нестационарными случайными процессами, их характеристики необходимо получать путем последовательного применения операций усреднения по множеству и по времени. Для определения корреляционной функции и спектральной плотности используют обобщенные соотношения Хинчина-Винера (3.3), (3.4).

Классификационная матрица видов модуляции (табл. 1) охватывает все изученные виды непрерывной, дискретной и цифровой модуляции и позволяет разрабатывать новые виды, полезные для развития теории информации и передачи сигналов.

2. Для анализа корреляционных и спектральных характеристик модулированных сигналов необходимо знать корреляционные и спектральные характеристики модулирующих сигналов и переносчиков, а также структуру и параметры оператора модуляции. Спектральную плотность модулированного сигнала получают из соотношения Хинчина — Винера (3.3) по его корреляционной

функции. Упростить определение корреляционной функции позволяет представление модулированного сигнала в виде узкополосного процесса. Корреляционная функция балансно-модулированного высокочастотного колебания равна произведению корреляционной функции огибающей на корреляционную функцию высокочастотного сигнала.

3. В роли переносчиков часто используют узкополосные и широкополосные случайные процессы. Широкое применение получили дискретные шумоподобные сигналы, построенные на основе линейных рекуррентных последовательностей. Например, использование бинарных *М*-последовательностей Хаффмена позволяет получить сигналы-переносчики, обладающие рядом полезных свойств. Шумоподобные сигналы наряду с известными параметрическими видами модуляции можно подвергать структурной модуляции (модуляции по форме).

4. Спектр узкополосного аналитического случайного сигнала, модулированного по амплитуде случайным сигналом, является непрерывным. Он сосредоточен в области средней частоты случайной несущей и имеет две боковые полосы, отстоящие от этой частоты на среднюю частоту спектра полезного сигнала.

5. Анализ АМн и АИМ сигналов позволяет выявить все основные особенности анализа модулированных колебаний при дискретной и импульсной модуляциях. Спектр гармонического колебания, амплитудно-манипулированного случайным телеграфным сигналом, образуется в результате переноса спектра телеграфного сигнала на частоту несущей и «наложения» этого спектра на спектральную линию гармонического колебания. Спектр АИМ сигнала является упорядоченным набором спектров амплитудно-модулированных колебаний, в которых роль несущих выполняют гармоники частоты следования видеоимпульсов.

6. Цифровые виды модуляции применяют для передачи непрерывных сообщений дискретными методами. При цифровой модуляции производят дискретизацию во времени и квантование по уровню непрерывного сигнала и полученные квантованные значения сигнала кодируют. Основными видами цифровой модуляции являются импульсно-кодовая модуляция (ИКМ), дельта-модуляция (ДМ) и комбинированные виды (ИКМ-ДМ). Достоинствами систем с цифровым способом передачи сигналов являются: высокая помехоустойчивость, возможность регенерации сигналов в узлах связи, низкая чувствительность к нелинейным искажениям; интеграция способов обработки, коммутации и управления сигналами, унификация функциональных узлов на основе микросхемотехники и цифровой вычислительной техники. Однако применение цифровых методов ведет к расширению абонируемой полосы частот каналов, к необходимости точной синхронизации сигналов и построения аппаратуры для регенерации.

В настоящее время системы с ИКМ широко применяют в телефонии и спутниковой связи. Область использования цифровых методов модуляции постоянно расширяется.

Глава 4

КАНАЛЫ ПЕРЕДАЧИ ИНФОРМАЦИИ

4.1. АНАЛИЗ НЕПРЕРЫВНЫХ КАНАЛОВ

Методы и модели анализа непрерывных каналов разрабатывают на основании изучения физических и статистических характеристик реальных каналов. Так как непрерывные каналы являются основной составной частью всех других каналов, результаты анализа непрерывных каналов широко используют для решения задач анализа и синтеза систем, сетей связи и других объектов информационной техники. Основными задачами анализа непрерывных каналов являются анализ линейных и нелинейных искажений сигналов в каналах и анализ влияния помех в каналах.

4.1.1. Анализ искажений сигналов. Для анализа искажений сигналов в каналах необходимо располагать сведениями о характеристиках входных сигналов, структуре и параметрах операторов преобразования сигналов в канале и изучать характеристики выходных сигналов. Характеристики входных сигналов определяют как характеристики модулированных сигналов (см. § 3.2—3.6). Структуру и параметры операторов преобразования сигналов в канале определяют на основе построения математических моделей каналов (см. п. 4.1.3). Прохождение сигналов через каналы и характеристики выходных сигналов обычно изучают методами теории радиотехнических цепей [7] и статистической радиотехники [1—3, 15].

При строгом рассмотрении *реальные непрерывные каналы* являются *нелинейными инерционными стохастическими системами* [1—3]. В них реакция на выходе не может предшествовать воздействию на входе, поэтому такие системы часто называют *динамическими*. Анализ таких систем представляет сложную задачу. Ее решение еще более усложняется, когда в роли входных воздействий выступают случайные модулированные сигналы. Для приближенного решения задач анализа искажений непрерывный канал, как уже отмечалось в § 1.3, удобно рассматривать как последовательное соединение линейной инерционной системы и нелинейной, но безынерционной системы. На рис. 4.1 показана структурная схема непрерывного канала без помех, где линейная инерционная система представлена полосовым фильтром *L1*, а нелинейная безынерционная система — нелинейным преобразовате-

лем $L2$. В статистической радиотехнике [15] показано, как анализируют прохождение случайных сигналов через такие системы.

Линейные искажения сигналов появляются в линейном инерционном четырехполюснике с постоянными параметрами из-за наличия в нем реактивных элементов. При линейных искажениях нарушаются существующие частотные и фазовые соотношения между отдельными составляющими сигнала и форма сигналов. Для отсутствия искажений необходимо, чтобы модуль коэффициента передачи и время запаздывания для всех составляющих были одинаковы. Нелинейными называют искажения сигналов, которые возникают в нелинейных безынерционных четырехполюсниках с постоянными параметрами из-за нелинейности характе-

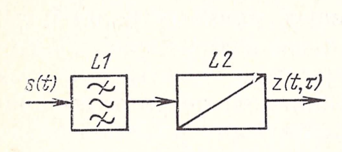


Рис. 4.1. Эквивалентная схема непрерывного канала без помех

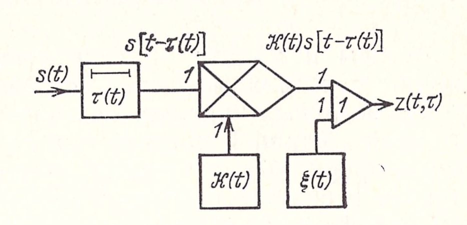


Рис. 4.2. Эквивалентная схема непрерывного канала с помехами

ристик активных элементов: ламп, транзисторов и др. В результате нелинейных искажений спектры сигналов расширяются, в них появляются дополнительные компоненты, растут уровни взаимных помех в каналах.

4.1.2. Помехи в непрерывных каналах. Для рассмотрения помех в непрерывных каналах выходной сигнал представляют в виде

$$z(t,\ \tau)=\mathscr{K}(t)\,s[t-\tau(t)]+\xi(t), \qquad (4.1)$$

где $s(t)$ — входной сигнал; $\mathscr{K}(t)$ и $\xi(t)$ — соответственно мультипликативная и аддитивная помехи; $\tau(t)$ — задержка сигнала в канале. Структурная схема непрерывного канала с помехами показана на рис. 4.2.

Мультипликативные помехи обусловлены случайными изменениями коэффициента передачи канала из-за изменения характеристик среды, в которой распространяются сигналы, и коэффициентов усиления схем при изменении питающих напряжений, из-за замираний сигналов в результате интерференции и различного затухания сигналов при многолучевом распространении радиоволн [9]. Сущность физических явлений, вызывающих мультипликативные помехи, подробно рассмотрена в [1, 9]. Мультипликативные помехи бывают «медленные», когда

$$\Delta\tau_{\text{к}}/\Delta\tau_{\text{c}}\gg 1, \qquad (4.2)$$

и «быстрые», когда

$$\Delta\tau_к/\Delta\tau_с \ll 1,$$

где $\Delta\tau_к$ — интервал корреляции случайного процесса $\mathscr{K}(t)$, $\Delta\tau_с$ — интервал корреляции или длительность сигнала, если он рассматривается как детерминированный.

Если сигнал включает ряд спектральных компонент и $\Delta\tau_i$ — интервал корреляции или длительность i-й компоненты сигнала, то в зависимости от значения отношения $\Delta\tau_к/\Delta\tau_i$ различают общие и селективные мультипликативные помехи (замирания сигналов). Если

$$\Delta\tau_к/\Delta\tau_i = \text{const}, \quad i = 1, 2, \ldots, \tag{4.3}$$

то мультипликативную помеху называют общей. Если это отношение различно для различных компонент, то помеху называют селективной. Если случайный сигнал может быть представлен в виде тригонометрического ряда Фурье (2.45), то в роли $\Delta\tau_i$ выступает период i-й гармоники $T_i = 2\pi/\omega_i$.

Аддитивные помехи обусловлены флуктуационными явлениями, связанными с тепловыми процессами в проводах, резисторах, лампах, транзисторах и других элементах схем, наводками под действием атмосферных явлений (грозовые, разряды, космическое излучение, магнитные бури и т. п.) и индустриальных процессов (работа промышленных установок, линий электропередач, радиостанций, других линий связи и т. п.).

Аддитивные помехи делят на сосредоточенные и флуктуационные. Сосредоточенные аддитивные помехи отличаются сосредоточенностью энергии помехи в полосе частот (узкополосные помехи) или на отрезке времени (импульсные помехи). Узкополосные помехи в основном обусловлены действием посторонних источников сигналов — ширина спектра этих помех сравнима или значительно меньше ширины спектра полезных сигналов. Узкополосные помехи как помехи от соседних станций характерны для радиосвязи. Статистические свойства узкополосных помех носят такой же характер, как и у полезных сигналов. Борьба с узкополосными аддитивными помехами ведется методами повышения избирательности радиоприемных устройств и улучшения линейности характеристик усилителей (нелинейные преобразования помех приводят к расширению их спектра, что вызывает появление частотных компонент помехи в полосе прозрачности систем, отведенной для приема полезных сигналов).

Импульсные помехи — это случайные последовательности импульсов, создаваемые промышленными установками и атмосферными источниками сигналов. Эти помехи характеризуются широким энергетическим спектром. Ширина их спектра, как известно, обратно пропорциональна длительности импульсов. Энергия спектральных составляющих импульсных помех падает в области сверхнизких и сверхвысоких частот. Это является одной из причин все более широкого использования радиоволн метрового, дециметрового и сантиметрового диапазонов.

Понятие сосредоточенности энергии помехи относительно. Поэтому для определенности сосредоточенными аддитивными помехами следует считать те, для которых

$$\Delta\omega_2/\Delta\omega_1 \ll 1, \quad T_2/T_1 \ll 1, \qquad (4.4)$$

где $\Delta\omega_2$, T_2 — соответственно ширина спектра и длительность помехи; $\Delta\omega_1$, T_1 — ширина спектра и длительность сигнала. Первое соотношение в (4.4) определяет узкополосную помеху, второе — импульсную.

Флуктуационная аддитивная помеха характеризуется «размытостью» энергии спектра в широком диапазоне частот. Она обусловлена главным образом внутренними шумами элементов аппаратуры (тепловые шумы, дробовой эффект в электровакуумных приборах и т. п.). Средняя мощность теплового шума в полосе частот ΔF полезного сигнала определяется по формуле

$$\mathcal{P} = \sigma^2 = 4kT^\circ\Delta F, \qquad (4.5)$$

спектральная плотность

$$\Omega = \mathcal{P}/\Delta F = 4kT^\circ, \qquad (4.6)$$

где $k = 1{,}37 \cdot 10^{-23}$ Дж/град — постоянная Больцмана; T° — абсолютная температура; при $T^\circ = 290$ К $kT^\circ = 4 \cdot 10^{-2i}$ Вт/Гц. Спектральная плотность помехи на положительных частотах $\Omega_0 = 2\Omega$. Флуктуационную помеху из-за «внутренней» природы невозможно устранить, можно лишь учесть ее характеристики при синтезе такой оптимальной системы, в которой наличие флуктуационной помехи меньше всего сказывается на качестве передачи информации.

Математическими моделями сосредоточенных аддитивных помех являются узкополосные случайные сигналы и случайные последовательности импульсов. Математической моделью флуктуационной аддитивной помехи служит гауссовский белый шум (см. п. 2.4.4).

4.1.3. Модели непрерывных каналов. В настоящее время разработано большое количество моделей непрерывных каналов, различных по сложности математического описания, требуемым исходным данным и погрешностям описания реальных каналов. Наиболее распространены следующие модели: идеальный канал, гауссов канал, гауссов канал с неопределенной фазой, гауссов однолучевой канал с замираниями, гауссов многолучевой канал с замираниями и сосредоточенными аддитивными помехами. Для анализа реальных каналов в конкретных условиях обычно выбирают такую модель, которая приводит к не слишком трудоемким решениям задач и в то же время обладает погрешностями, допустимыми в инженерных расчетах.

Идеальный канал можно применять как модель реального непрерывного канала, если соблюдаются следующие условия: помехи любого вида отсутствуют, оператор $L = L_1L_2$ преобразования сигналов в канале является детерминированным (см. рис. 4.1), мощность и полоса сигналов ограничены. Для анализа выходных

сигналов с помощью этой модели необходимо знать характеристики входных сигналов и операторов L_1, L_2. Модель идеального канала слабо отражает реальные условия, ее применяют чаще всего для анализа линейных и нелинейных искажений модулированных сигналов в многоканальных системах проводной связи.

Гауссовский канал. Основные допущения при построении этой модели следующие: коэффициент передачи и время задержки сигналов в канале не зависят от времени и являются детерминированными величинами, известными в месте приема сигналов; в канале действует аддитивная флуктуационная помеха — гауссовский белый шум (гауссовский процесс).

Если на вход гауссового канала поступает узкополосный сигнал, то выходной сигнал можно представить в виде

$$z(t, \mathscr{K}, \varphi) = \mathscr{K} \cos \varphi s(t) + \mathscr{K} \sin \varphi \widehat{s}(t) + \xi(t), \tag{4.7}$$

где $s(t)$, $\hat{s}(t)$ — квадратурные составляющие входного сигнала; $\mathscr{K}$ — коэффициент передачи канала как функция времени; $\varphi = \omega_0 \tau$; ω_0 — средняя частота входного сигнала; τ — время задержки сигнала в канале; $\xi(t)$ — гауссовский белый шум. Если на вход гауссова канала поступает широкополосный сигнал, для i-й компоненты которого коэффициент передачи канала равен $\mathscr{K}_i$, а фазовый сдвиг φ_i, то выходной сигнал

$$z(t) = \sum_{i=1}^{n} [\mathscr{K}_i \cos \varphi_i s_i(t) + \mathscr{K}_i \sin \varphi_i \widehat{s}_i(t)] + \xi(t), \tag{4.8}$$

где $\varphi_i = \omega_i \tau_i$; ω_i — средняя частота i-й компоненты; τ_i — время задержки i-й компоненты; n — число компонент. Из сравнения (4.7) и (4.8) следует, что входной сигнал может рассматриваться как узкополосный, если амплитудные и фазовые искажения отсутствуют, и $\mathscr{K}_i = \mathscr{K}$, $\varphi_i = \varphi$, $i = \overline{1, n}$. Для анализа сигналов на выходе гауссовых каналов необходимо знать характеристики входных сигналов, значения $\mathscr{K}_i$ и φ_i, а также спектр Ω помехи $\xi(t)$.

Гауссов канал применяют как модель реальных каналов проводной связи и однолучевых каналов без замираний или с медленными замираниями, когда можно надежно измерить $\mathscr{K}_i$ и φ_i. Эта модель позволяет анализировать амплитудные и фазовые искажения сигналов и влияние флуктуационной помехи.

Гауссовский канал с неопределенной фазой сигнала. В этой модели время задержки сигнала в канале рассматривают как случайную величину, поэтому фаза в (4.7) также случайна. Для анализа выходных сигналов канала необходимо знать закон распределения времени задержки или фазы сигнала.

Гауссовский однолучевой канал с замираниями. В этой модели $\mathscr{K}$ и φ рассматривают в (4.7) как случайные величины или процессы. Если $\mathscr{K}(t)$ и $\varphi(t)$ — случайные процессы, спектр выходного сигнала канала шире спектра входного даже при отсутствии помехи из-за паразитных амплитудной и фазовой модуляций.

Введем в (4.7) следующие обозначения для квадратурных компонент:

$$X(t) = \mathscr{K}(t)\cos\varphi(t), \quad Y(t) = \mathscr{K}(t)\sin\varphi(t). \qquad (4.9)$$

Для реальных каналов измеряют следующие характеристики этих процессов: математические ожидания $M[X(t)] = m_1$, $M[Y(t)] = m_2$, дисперсии $\mathscr{D}[X(t)] = \sigma^2{}_1$, $\mathscr{D}[Y(t)] = \sigma^2{}_2$, корреляционные функции $K_1(\tau)$, $K_2(\tau)$. В зависимости от измеренных значений характеристик, различают обобщенную гауссовскую модель, обобщенную релеевскую модель и релеевскую модель однолучевого канала с замираниями.

Для обобщенной гауссовской модели канала $K_1(\tau) = K_2(\tau) = K(\tau)$ процессы $X(t)$, $Y(t)$ являются некоррелированными, гауссовскими и стационарными. Для анализа каналов с помощью этой модели необходимо знать характеристики входных сигналов m_1, m_2, $\sigma^2{}_1$, $\sigma^2{}_2$, $K(\tau)$, спектр помехи Ω. Основная особенность анализа выходных сигналов заключается в том, что они являются случайными нестационарными процессами, корреляционные функции и спектральные плотности которых необходимо находить методами, изложенными в § 3.2.

В обобщенной релеевской модели канала $\sigma^2{}_1 = \sigma^2{}_2 = \sigma^2$, распределение модуля коэффициента передачи канала

$$\mathscr{K} = \sqrt{X^2 + Y^2} \qquad (4.10)$$

является обобщенным релеевским распределением (2.86), где роль величины A играет $\mathscr{K}$, а роль величины $a^2{}_0$ — параметр $q^2 = (m^2{}_1 + m^2{}_2)/2\sigma^2$. Распределение фазы

$$\varphi = \mathrm{arctg}\,(Y/X) \qquad (4.11)$$

имеет вид (2.87).

В релеевской модели канала $\sigma^2{}_1 = \sigma^2{}_2 = \sigma^2{}_\text{к}/2$, $m_1 = m_2 = 0$, поэтому распределение величины (4.10) является распределением Релея (2.78), а распределение фазы равномерное (2.79). Следовательно, обобщенная гауссовская модель однолучевого канала с замираниями является наиболее общей, частными видами этой модели служат обобщенная релеевская модель и релеевская модель.

Рассмотренные модели однолучевого канала с замираниями достаточно хорошо описывают свойства радиоканалов различных диапазонов и проводных каналов со случайными, в том числе и переменными параметрами.

Гауссов многолучевой канал с замираниями. Эта модель описывает радиоканалы, распространение сигналов от передатчика к приемнику в которых происходит по различным «каналам» — путям. Длительности прохождения сигналов и коэффициенты передачи различных «каналов» являются неодинаковыми и случайными. Принимаемый сигнал образуется в результате интерференции сигналов, пришедших по различным путям. Он описывается соотношением (4.8), в котором $s_i(t)$ и $\hat{s}_i(t)$ — квадратурные составляющие передаваемого сигнала, прошедшие по i-му пути, $\mathscr{K}_i$

и φ_i — коэффициент передачи и фаза для i-го пути, n — число путей распространения радиоволн.

В общем случае частотная и фазовая характеристики канала зависят от времени и частоты, поэтому модель коэффициента передачи канала

$$\mathcal{K}(\omega, t) = \sqrt{\left[\sum_{i=1}^{n} \mathcal{K}_i(\omega, t)\cos\varphi_i(\omega, t)\right]^2 + \left[\sum_{i=1}^{n} \mathcal{K}_i(\omega, t)\sin\varphi_i(\omega, t)\right]^2}, \tag{4.12}$$

фазовая характеристика канала

$$\varphi(\omega, t) = \operatorname{arctg} \frac{\sum_{i=1}^{n} \mathcal{K}_i(\omega, t)\sin\varphi_i(\omega, t)}{\sum_{i=1}^{n} \mathcal{K}_i(\omega, t)\cos\varphi_i(\omega, t)}. \tag{4.13}$$

Следовательно, для описания многолучевых каналов с замираниями необходимо задавать в n раз больше статистических характеристик по сравнению с однолучевыми. Из-за этого задачи анализа многолучевых каналов являются значительно более сложными. В то же время модель многолучевого канала с замираниями является одной из наиболее общих и пригодна для описания свойств большинства радиоканалов и проводных каналов.

Гауссовский многолучевой канал с замираниями и аддитивными сосредоточенными помехами. В этой модели наряду с флуктуационной помехой учитывают и различного вида сосредоточенные помехи. Она является наиболее общей и достаточно полно отражает свойства многих реальных каналов. Однако ее использование порождает сложность и трудоемкость решения задач анализа, а также необходимость сбора и обработки большого объема исходных статистических данных.

В дальнейшем для решения задач анализа непрерывных и дискретных каналов используются, как правило, модель гауссовского канала и модель гауссовского однолучевого канала с замираниями.

Контрольные вопросы

1. Нарисуйте и поясните структурную схему непрерывного канала без помех.
2. Какие искажения сигналов наблюдаются в непрерывных каналах?
3. Нарисуйте и поясните структурную схему непрерывного канала с помехами.
4. Какие помехи наблюдаются в непрерывных каналах?
5. Какие модели применяют для описания непрерывных каналов?
6. Опишите модель гауссовского канала, гауссовского однолучевого канала с замираниями.

4.2. АНАЛИЗ ДИСКРЕТНО-НЕПРЕРЫВНЫХ КАНАЛОВ

Чтобы показать особенности анализа дискретно-непрерывных и непрерывно-дискретных каналов, рассмотрим, как выполняется анализ дискретно-непрерывных каналов. Для определенности выберем дискретно-непрерывный канал, который образуется устройствами тракта «выход кодера — вход демодулятора $D_{\text{м}2}$» (рис. 1.1). На вход этого канала поступают дискретные кодовые сигналы $b_{\text{к}1}$, а с выхода снимаются непрерывные сигналы $s_2(t)$.

Для построения математической модели и анализа дискретно-непрерывного канала необходимо знать алфавит кодовых сигналов b_{k1}, $k=\overline{1,m}$, вероятности появления кодовых сигналов $P(b_{k1})$, полосу пропускания $\Delta F_{\text{к}}$ непрерывного канала, на котором построен рассматриваемый канал, априорную условную плотность $f_1(s|b_{k1})$ распределения вероятности появления сигнала $s_2(t)$ при условии, что передавался сигнал b_{k1}. Первые две характеристики определяют свойства кодера как источника дискретных сообщений. Полоса пропускания непрерывного канала определяет допустимую скорость W передачи кодовых сигналов и их искажения. Четвертая характеристика определяется видом мультипликативных и аддитивных помех и искажений в непрерывном канале.

Результатом анализа дискретно-непрерывного канала является определение апостериорной вероятности $P(b_{k1}|s)$ того, что при полученном сигнале $s_2(t)$ передавался сигнал b_{k1}. Апостериорная вероятность рассчитывается по формуле Байеса:

$$P(b_{k1}|s)=\frac{P(b_{k1})\,f_1(s|b_{k1})}{\sum\limits_{k=1}^{m}P(b_{k1})\,f_1(s|b_{k1})}. \quad (4.14)$$

Если решающая схема демодулятора $D_{\text{м}2}$ реализует алгоритм определения максимума апостериорной вероятности:

$$\max\,[P(b_{k1}|s)]=P(b_{i1}|s), \quad k=\overline{1,\ m}, \quad (4.15)$$

то на выходе демодулятора появляется сигнал b_{i1}, апостериорная вероятность появления которого $P(b_{i1}|s)$ больше всех остальных.

Скорость передачи сигналов W выбирают из условия отсутствия межсимвольных искажений (искажений соседних импульсов из-за переходных процессов в канале). Ширина спектра сигнала $s_2(t)$ на выходе непрерывного канала не может превышать $\Delta F_{\text{к}}$. Такой сигнал полностью определяется отсчетами, следующими с частотой дискретизации $2\Delta F_{\text{к}}$. Поэтому скорость передачи сигналов должна удовлетворять условию

$$W\leqslant 2\Delta F_{\text{к}}. \quad (4.16)$$

В дискретных каналах скорость передачи сигналов может быть и больше чем $2\Delta F_{\text{к}}$ (величину $2\Delta F_{\text{к}}$ называют пределом Найквиста). Однако в этом случае появляются межсимвольные искажения и нельзя проводить независимую обработку отдельных сигналов.

Необходимо применять достаточно сложную аппаратуру компенсации искажений. Из-за наличия помех полностью устранить искажения невозможно, что в конечном итоге приводит к ошибкам решающей схемы $D_{м2}$. Знак равенства в (4.16) достигается при независимой обработке отдельных кодовых сигналов в том случае, когда импульсная переходная характеристика непрерывного канала совпадает с функцией отсчета, т. е. когда канал является идеальным фильтром. Рассмотрение моделей непрерывных каналов показывает, что это условие, как правило, не выполняется. Поэтому в реальных каналах $W \leqslant 2\Delta F_к$.

Характер условной плотности $f_1(s|b_{k1})$ полностью определяет свойства дискретно-непрерывного канала. Если для любых сочетаний $s_2(t)$ и b_{k1}

$$f_1(s|b_{k1};\ t) = f_1(s|b_{k1}), \qquad (4.17)$$

то такой канал называют *стационарным*. Если условие (4.17) не выполняется, канал является нестационарным. Следовательно, для стационарного дискретно-непрерывного канала условная плотность $f_1(s|b_{k1})$ не зависит от времени.

Если справедливо условие

$$f_1(s|b_{k1},\ b_{k-1,1},\ b_{k-2,1},\ \ldots,\ b_{k-l,1}) = f_1(s|b_{k1}), \qquad (4.18)$$

то такой канал называют *каналом без памяти*. В противном случае говорят, что канал обладает памятью. Выполнение условий (4.17), (4.18) определяется свойствами непрерывного канала, на основе которого построен дискретно-непрерывный канал. Если непрерывный канал является гауссовым, то условия (4.17), (4.18) выполняются и дискретно-непрерывный канал является *стационарным и без памяти*.

Реальные дискретно-непрерывные каналы обычно являются нестационарными и с памятью. Несмотря на это, модель дискретно-непрерывного стационарного канала без памяти находит широкое применение благодаря простоте анализа и получения исходных данных.

Контрольные вопросы

1. Какие устройства входят в дискретно-непрерывный канал?
2. Какие исходные данные требуются для анализа дискретно-непрерывных каналов?
3. Что является основным результатом анализа дискретно-непрерывного канала?
4. Как определяют скорость передачи сигналов?
5. Запишите и поясните условия стационарности и отсутствия памяти дискретно-непрерывного канала.

4.3. АНАЛИЗ ДИСКРЕТНЫХ КАНАЛОВ

Для анализа дискретных каналов разрабатывают и применяют специальные математические модели и методы. Рассмотрим основные из них и на примере двоичного канала покажем, как опре-

деляют характеристики дискретных каналов: условные вероятности появления ошибок, полные вероятности появления ошибки и правильного приема, вероятности появления различных символов на выходе дискретного канала и др.

4.3.1. Модели дискретных каналов. Дискретный канал образуют устройства тракта «вход кодера — выход декодера» (рис. 1.1). На вход канала поступают символы a_{k1}, а с выхода — a_{k2}. Математическая модель дискретного канала определена, если известны следующие характеристики: алфавит и априорные вероятности $P(a_{k1})$ появления символов a_{k1} сообщений ($k=\overline{1,m_1}$, m_1 — объем алфавита); скорость передачи символов W_1; алфавит символов a_{i2} копии сообщений, $i=\overline{1,m_2}$, m_2 — объем алфавита; априорная условная вероятность $P(a_{i2}|a_{k1})$ появления символа a_{i2} при условии, что был передан a_{k1}.

Как и для дискретно-непрерывного канала, первые две характеристики определяются свойствами источника сообщений и полосой пропускания непрерывного канала. Объем выходного алфавита m_2 определяется способом построения системы передачи информации. Условная вероятность $P(a_{i2}|a_{k1})$ определяется в основном свойствами непрерывного канала и его характеристиками. Если в системе используют канал обратной связи и «стирание» символов, то $m_2>m_1$. Стирание символов вводят тогда, когда из-за искажений и помех не ясно, какой символ передавался. Решающее устройство декодера $D_к$ выдает символ стирания θ, если символ a_{i2} настолько отличается от символов источника сообщений, что его нельзя с большой вероятностью отождествить ни с одним из передаваемых. Стирание символов позволяет уменьшить вероятность появления ошибки, но приводит к уменьшению и вероятности правильного приема. Определены условия, при которых стирание символов целесообразно. Обычно вводят один символ стирания.

Результатом анализа дискретного канала является определение *апостериорной условной вероятности* $P(a_{k1}|a_{i2})$ того, что при полученном символе a_{i2} передавался символ a_{k1}. С помощью этих апостериорных вероятностей и априорных вероятностей $P(a_{k1})$ рассчитывают полную вероятность появления ошибки в канале, полную вероятность правильного приема, вероятность появления символов на выходе канала, информационные характеристики дискретного канала (скорость передачи информации, пропускную способность, количество принятой информации и др. (см. § 5.2)).

Апостериорная вероятность рассчитывается по формуле Байеса

$$P(a_{k1}|a_{i2})=\frac{P(a_{k1})\,F(a_{i2}|a_{k1})}{\sum\limits_{k=1}^{m_1}P(a_{k1})\,P(a_{i2}|a_{k1})}. \qquad (4.19)$$

Если решающая схема декодера $D_к$ реализует алгоритм определения максимума апостериорной вероятности:

$$\max_k\,[P(a_{k1}|a_{i2})]=P(a_{j1}|a_{i2}), \qquad (4.20)$$

то на выходе декодера появляется символ a_{j1}, апостериорная вероятность появления которого $P(a_{j1}|a_{i2})$ больше всех остальных.

Характер условных вероятностей $P(a_{i2}|a_{k1})$ полностью определяет свойства дискретного канала. Если для любых сочетаний a_{i2} и a_{k1} эта вероятность не зависит от момента времени t взятия отсчета, т. е.

$$P(a_{i2}|a_{k1};\ t)=P(a_{i2}|a_{k1}), \tag{4.21}$$

то канал называют *однородным*. Если условие (4.21) не выполняется, канал является неоднородным. Если справедливо условие

$$P(a_{i2}|a_{k1},\ a_{k-1,1},\ \ldots,\ a_{k-l,1})=P(a_{i2}|a_{k1}), \tag{4.22}$$

то канал называют *каналом без памяти*. Если условие (4.22) не выполняется, канал обладает памятью на l символов. Выполнение условий (4.21) и (4.22) зависит от того, на каком непрерывном канале построен дискретный канал. Например, если непрерывный канал является гауссовым, то условия (4.21), (4.22) выполняются и построенный на нем дискретный канал является *однородным и без памяти*.

Реальные дискретные каналы являются неоднородными и с памятью. Это обусловлено следующими причинами: искажением сигналов и влиянием помех в непрерывном канале (см. § 4.1), задержкой во времени выходной последовательности сигналов по отношению к входной, нарушением тактовой синхронизации передаваемых и принимаемых импульсов (см. § 6.6), ошибками решающих схем (см. п. 4.3.3). Однако модель дискретного однородного канала без памяти как модель первого приближения нашла широкое применение. Она позволяет упростить методы анализа и получение исходных данных.

4.3.2. Методы определения характеристик дискретных каналов. Для математического описания дискретных однородных каналов без памяти необходимо использовать матрицы типа

$$p_{ik}=\begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1m} \\ p_{21} & p_{22} & \cdots & p_{2m} \\ \vdots & \vdots & & \vdots \\ p_{m1} & p_{m2} & \cdots & p_{mm} \end{pmatrix}, \tag{4.23}$$

элементами которых являются условные вероятности $p_{ik}=$ $=P(a_{i2}|a_{k1})$ (см. п. 4.3.1). Совместно с априорными вероятностями $P(a_{k1})$ эти вероятности p_{ik} перехода i-го символа в k-й полностью определяют вероятностные характеристики дискретных каналов. Математическим аппаратом, который позволяет исследовать дискретные каналы, является теория марковских цепей [15]. Она предназначена для описания случайных дискретных последовательностей. Рассмотрим те элементы этой теории, которые используются в дальнейшем.

Если выполнить дискретизацию случайного процесса $X(t)$ с интервалом $\Delta t=1/2F$ (см. § 2.2), то значения случайного процесса $X_i=X(t_i)$, взятые в моменты времени t_i, образуют случайную

последовательность $\{X_i\}$. Если случайная последовательность получена дискретизацией стационарного и эргодического процесса, она также обладает этими свойствами. Числовые характеристики такой последовательности получают использованием операций усреднения по множеству и по времени (см. § 2.3). Оценка математического ожидания последовательности

$$m = M[X_i] \approx \langle X_i \rangle = \frac{1}{n} \sum_{k=1}^{n} x_{ik}, \qquad (4.24)$$

где при усреднении по множеству n — количество реализаций, измеренных в один момент времени t_i; x_{ik} — k-е значение случайной величины X_i; при усреднении по времени n — количество моментов времени, рассматриваемых для одной реализации.

Если все значения X_i стационарной последовательности непрерывны и независимы, то полной характеристикой является одномерная плотность распределения $f(x_i)$. Плотности распределения большей размерности определяют как произведение одномерных плотностей. Если X_i являются дискретными независимыми символами, что имеет место при определенных условиях передачи дискретных сообщений, полной характеристикой является распределение вероятностей p_i появления символа X_i, $i = \overline{1, n}$. Так как X_i образуют полную группу событий, то

$$\sum_{i=1}^{n} p_i = 1. \qquad (4.25)$$

Равенство (4.25) называют *условием нормировки*.

Если символы последовательности взаимозависимы (коррелированы), помимо вероятностей появления отдельных символов, необходимо задавать условные вероятности $P(X_i | X_{i-1}, X_{i-2}, \ldots$ $\ldots, X_{i-l})$ появления в последовательности символа X_i при условии, что перед ним появилась группа символов $X_{i-1}, X_{i-2}, \ldots$ $\ldots, X_{i-l}$. Последовательности, в которых существуют статистические связи между символами, называют цепями Маркова или *марковскими цепями*. Если статистическая связь существует только между двумя символами i-м и $(i-1)$-м, то марковскую цепь называют простой, ее поведение полностью описывается матрицей (4.23) при заданных начальных вероятностях $P(a_{k1}) = p_k$. Для эргодической марковской цепи вероятности p_j появления символов X_j в установившемся режиме находят из системы алгебраических уравнений

$$p_j = \sum_{i=1}^{m} p_i p_{ij}, \quad j = \overline{1, m}, \qquad (4.26)$$

с использованием условия нормировки (4.25).

Для математического описания дискретных однородных каналов без памяти используют методы теории марковских простых однородных цепей.

4.3.3. Двоичный канал. Используя результаты п. 4.3.1, 4.3.2, найдем вероятностные характеристики *двоичного дискретного однородного канала без памяти.* В этом случае $m_1 = m_2 = 2$. Для простоты записи обозначим $b_{11} = b_{12} = b_1$, $b_{21} = b_{22} = b_2$, $p_{11} = P(b_1|b_1)$, $p_{12} = P(b_1|b_2)$, $p_{21} = P(b_2|b_1)$, $p_{22} = P(b_2|b_2)$. Вероятности p_{11}, p_{22} — это условные вероятности правильного приема символов b_1, b_2, а p_{12} и p_{21} — это условные вероятности появления ошибок. Граф преобразования символов при передаче по двоичному каналу показан на рис. 4.3.

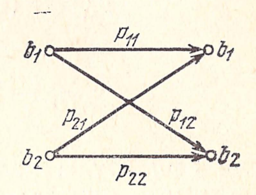


Рис. 4.3. Граф преобразования символов в дискретном двоичном канале

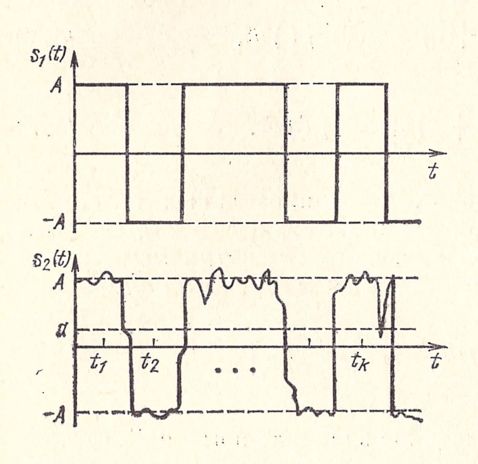


Рис. 4.4. Реализации сигналов $s_1(t)$ и $s_2(t)$

Рассмотрим работу решающей схемы $D_{м2}$. Реализации сигналов $s_1(t)$ на выходе первого модулятора и $s_2(t)$ на входе демодулятора $D_{м2}$ (см. рис. 1.1) показаны на рис. 4.4. Положительные импульсы соответствуют передаче символа b_1, отрицательные — передаче b_2. Можно заметить, что прохождение сигнала через канал привело к изменению его формы.

Если $W < 2\Delta F$, искажения сигналов в канале отсутствуют и непрерывный канал является гауссовым, то изменение формы сигнала обусловлено лишь действием флуктуационной помехи $\xi(t)$. Сигнал на входе решающей схемы можно представить в виде $s_2(t) = s_1(t) + \xi(t)$.

На основании отсчетов напряжения принятого сигнала $s_2(t)$ в моменты времени $t_1, t_2, \ldots, t_k, \ldots$ решающая схема демодулятора $D_{м2}$ должна определить: был принят импульс с амплитудой $+A$ или импульс с амплитудой $-A$. Так как $|A|$ является детерминированной величиной, то распределение суммы $|A| + \xi(t_k)$ полностью определяется одномерным распределением помехи $f(\xi)$.

Вероятности ошибок и правильного приема определяются не только характеристиками помех, но и порогом a принятия решения. Если $s_2(t_k) < a$, то принимается решение о том, что пришел

отрицательный импульс. Правильные решения принимаются тогда, когда выполняются следующие неравенства

$$A+\xi_k>a,\ \ \xi_k>-A+a, \tag{4.27}$$

$$-A+\xi_k<a,\ \ \xi_k<A+a. \tag{4.28}$$

Ошибки происходят тогда, когда неравенства (4.27), (4.28) не выполняются из-за выбросов, обусловленных помехой. Условные вероятности ошибок — это вероятности выполнения противоположных неравенств, поэтому

$$p_{21}=P(\xi<-A+a)=\int\limits_{-\infty}^{-A+a} f(\xi)\,d\xi, \tag{4.29}$$

$$p_{12}=P(\xi>A+a)=\int\limits_{A+a}^{\infty} f(\xi)\,d\xi. \tag{4.30}$$

Если амплитуды положительных и отрицательных импульсов передаваемого сигнала одинаковы, удобно взять $a=0$. В этом случае $p_{12}=p_{21}=p_0$. Такой канал называют *симметричным*. *Условная вероятность появления ошибки в симметричном канале*

$$p_0=\int\limits_{A}^{\infty} f(\xi)\,d\xi=\int\limits_{-\infty}^{-A} f(\xi)\,d\xi=\frac{1}{2}\,[1-F_{\text{К}}(\sqrt{h})], \tag{4.31}$$

где $\sqrt{h}=A/\sigma$; σ — среднеквадратическое значение помехи; $F_{\text{К}}(\sqrt{h})$ — функция Крампа (2.88).

Безусловную вероятность ошибки Q определим по формуле полной вероятности

$$Q=P(b_1)\,p_{21}+P(b_2)\,p_{12}=p_0. \tag{4.32}$$

Из-за симметрии двоичного канала полная вероятность ошибки совпадает с условной вероятностью. Это удобное свойство симметричного канала, так как значение p_0 (одного параметра) полностью определяет свойства двоичного однородного симметричного канала без памяти. Полная вероятность правильного приема сигналов

$$P=P(b_1)\,p_{11}+P(b_2)\,p_{22}=1-p_0, \tag{4.33}$$

так как $p_{11}=1-p_{12}$, а $p_{22}=1-p_{21}$.

Реальный дискретный канал можно рассматривать как функциональный преобразователь распределения вероятностей появления символов входного алфавита в распределение вероятностей появления символов выходного алфавита. Идеальный дискретный канал не является преобразователем, поскольку оставляет распределение символов неизменным и оригиналы и копии дискретных сообщений совпадают.

4.3.4. Вероятность появления в кодовой комбинации q ошибок. Так как символы дискретных сообщений кодируют кодовыми

комбинациями, которые включают n элементарных кодовых сигналов, представляет интерес определение вероятности того, что в кодовой комбинации будет q ошибочно принятых элементарных сигналов. Величину q называют кратностью ошибок. Если все элементарные сигналы в кодовой комбинации независимы, эта вероятность определяется биномиальным распределением и формулой Бернулли:

$$P_n(p_0,\ q)=C^q{}_n p^q{}_0(1-p_0)^{n-q}, \tag{4.34}$$

где $C^q{}_n=n!/q!(n-q)!$ — число сочетаний; p_0 — вероятность появления ошибки при передаче одного элементарного сигнала. (Определение этой вероятности с учетом группирования ошибок дано в § 8.1.)

Среднее число ошибок

$$M[q]=\sum_{q=0}^{n} qP_n(p_0,\ q)=np_0. \tag{4.35}$$

Если $p_0 \ll 1$, что справедливо для реальных каналов, максимальной является вероятность $P_n(p_0,\ 0)$ того, что ошибок не будет. С ростом q функция $P_n(p_0,\ q)$ монотонно убывает. Поэтому ошибки большой кратности (когда $q>1$) встречаются реже. Этот вывод справедлив для однородных каналов без памяти при условии, что

$$np_0 \ll 1. \tag{4.36}$$

Поэтому в первую очередь обращают внимание на обнаружение и исправление ошибок малой кратности.

Контрольные вопросы

1. Какие характеристики необходимо знать для построения модели дискретного канала?
2. Что определяют в результате анализа дискретного канала?
3. Какие дискретные каналы называют однородными и без памяти?
4. Какой математический аппарат применяют для определения характеристик дискретных каналов?
5. По каким причинам реальные каналы отличаются от идеального?
6. Какие характеристики определяют для двоичного канала?
7. Как определить вероятность появления нескольких ошибок в одной кодовой комбинации?

4.4. ПРОХОЖДЕНИЕ СИГНАЛОВ ЧЕРЕЗ КАНАЛЫ

Для решения задачи прохождения сигналов через реальные каналы в общей постановке необходимо изучать прохождение случайных сигналов через нелинейные стохастические инерционные нестационарные системы. Работа таких систем описывается нелинейными дифференциальными уравнениями со случайными переменными коэффициентами и случайной правой частью. Поэтому решение таких задач является сложным, для многих реальных каналов оно является предметом современных научных ис-

следований. Характерные особенности задач анализа прохождения случайных сигналов через каналы обычно рассматривают с помощью более простых приближенных моделей каналов. Например, как уже отмечалось, для анализа линейных и нелинейных искажений сигналов канал рассматривают как последовательное соединение линейной инерционной системы и нелинейной безынерционной системы. Прохождение сигналов через такие системы изучают в статистической радиотехнике [15] и теории радиотехнических цепей [7].

Для систематизации представлений и иллюстрации основных особенностей анализа рассмотрим прохождение случайных стационарных сигналов через линейные инерционные системы с постоянными параметрами. Задачи анализа прохождения сигналов через нестационарные системы рассматриваются в § 6.6, где изучается влияние замираний сигналов на помехоустойчивость.

Напомним, что линейной называют систему, подчиняющуюся принципу суперпозиции. Примерами линейных инерционных преобразований являются такие операции, как усиление, фильтрация, дифференцирования, интегрирование, которые описывают линейным оператором (см. приложение). Полное описание работы линейной системы выполняют с помощью дифференциального уравнения с постоянными коэффициентами. Широкое распространение получили импульсные и частотные характеристики линейных систем, которые обычно применяют для анализа установившихся режимов [7]. Задача анализа прохождения случайных сигналов через линейную систему в полной мере решена тогда, когда по заданной n-мерной плотности распределения входного сигнала $X(t)$ и импульсной $g(t)$ (или коэффициенту передачи $\mathscr{K}(i\omega)$) характеристике линейной системы найдена n-мерная плотность распределения выходного сигнала $Y(t)$. Если входной процесс является гауссовским, то распределение выходного процесса остается гауссовским. В тех случаях, когда распределение входного сигнала отличается от гауссовского, определение распределения выходного процесса является достаточно сложной задачей, решаемой с помощью специальных приемов [7].

В линейных системах прохождение сигнала и аддитивной помехи можно рассматривать отдельно, а затем просуммировать полученные отклики. Прохождение детерминированных (регулярных) сигналов через линейные системы детально рассматривается в курсе «Радиотехнические цепи» [7]. Основное внимание мы будем уделять прохождению случайных сигналов. В инженерной практике решение задач анализа выходных распределений во многих случаях упрощается в результате того, что в узкополосных линейных системах имеет место «нормализация» выходных распределений [15]. Эффект нормализации проявляется в том, что вне зависимости от характера входного распределения выходное распределение тем больше приближается к нормальному, чем уже полоса пропускания системы по сравнению с шириной спектра входного случайного процесса.

Рассмотрим преобразование гауссовского стационарного сигнала $X(t)$ с корреляционной функцией $K_1(\tau)$ и спектральной плотностью $\Omega_1(\omega)$ в линейной инерционной системе с импульсной характеристикой $g(t)$ и коэффициентом передачи $\mathscr{K}(i\omega)$. Найдем корреляционную функцию $K_2(\tau)$ и спектральную плотность $\Omega_2(\omega)$ выходного процесса $Y(t)$. В роли входного сигнала будем рассматривать центрированный процесс $X(t)$, у которого $M[X(t)]=0$. Это вызвано тем, что прохождение постоянной (детерминированной) и переменной (случайной) составляющих через линейные системы удобно рассматривать отдельно, а характеристики выходного колебания получать методом суперпозиции.

На основании метода Дюамеля выходной сигнал

$$Y(t)=\int_0^\infty g(t)\,X(t-\tau)\,d\tau, \tag{4.37}$$

поэтому корреляционная функция

$$K_2(\tau)=M\left[\int_0^\infty g(\tau_1)\,X(t-\tau_1)\,d\tau_1\int_0^\infty g(\tau_2)\,X(t+\tau-\tau_2)\,d\tau_2\right]. \tag{4.38}$$

Операции интегрирования в (4.38) и усреднения по множеству можно менять местами, поэтому

$$K_2(\tau)=\int_0^\infty\int_0^\infty g(\tau_1)\,g(\tau_2)\,M\left[X(t-\tau_1)\,X(t+\tau-\tau_2)\right]d\tau_1\,d\tau_2.$$

Следовательно,

$$K_2(\tau)=\int_0^\infty\int_0^\infty g(\tau_1)\,g(\tau_2)\,K_1(\tau+\tau_1-\tau_2)\,d\tau_1\,d\tau_2. \tag{4.39}$$

Из (4.39) следует, что при прохождении через линейную инерционную стационарную систему процесс остается стационарным. Дисперсия выходного процесса

$$\mathscr{D}_2=K_2(0)=\int_0^\infty\int_0^\infty g(\tau_1)\,g(\tau_2)\,K_1(\tau_1-\tau_2)\,d\tau_1\,d\tau_2. \tag{4.40}$$

Таким образом, для определения дисперсии выходного сигнала недостаточно знать дисперсию входного сигнала и $g(t)$. Требуется знать корреляционную функцию $K_1(\tau)$ при $\tau\neq 0$.

Спектральную плотность выходного сигнала определим с помощью соотношения Хинчина—Винера (2.22):

$$\Omega_2(\omega)=\int_{-\infty}^{\infty}\left[\int_0^\infty\int_0^\infty g(\tau_1)\,g(\tau_2)\,K_1(\tau+\tau_1-\tau_2)\,d\tau_1\,d\tau_2\right]e^{-i\omega\tau}\,d\tau.$$

Учтем, что

$$\int_{-\infty}^{\infty} K_1(\tau+\tau_1-\tau_2)\, e^{-i\omega(\tau+\tau_1-\tau_2)} d\tau = \Omega_1(\omega),$$

$$\int_{0}^{\infty} g(\tau_1)\, e^{i\omega\tau_1} d\tau_1 = K(i\omega), \quad \int_{0}^{\infty} g(\tau_2)\, e^{-i\omega\tau_2} d\tau_2 = \mathscr{K}^*(i\omega),$$

где $\mathscr{K}^*(i\omega)$ — величина, комплексно-сопряженная с $\mathscr{K}(i\omega)$. Следовательно,

$$\Omega_2(\omega) = \Omega_1(\omega)\mathscr{K}(i\omega)\mathscr{K}^*(i\omega) = \Omega_1(\omega)\,|\mathscr{K}(i\omega)|^2, \tag{4.41}$$

где $|\mathscr{K}(i\omega)|$ — модуль коэффициента передачи.

Из формулы (4.41) следует, что фазовая характеристика линейной системы не влияет на корреляционную функцию и спектральную плотность выходного процесса. Поэтому по изменениям корреляционной функции и спектральной плотности процесса нельзя судить о фазовой характеристике системы.

Дисперсию выходного процесса можно определить и с помощью формулы (2.23):

$$\mathscr{D}_2 = \frac{1}{\pi}\int_{0}^{\infty} \Omega_1(\omega)\,|\mathscr{K}(i\omega)|^2\, d\omega. \tag{4.42}$$

Если выходной гауссовский процесс центрирован, $M[Y(t)]=0$ и дисперсия полностью определяет его одномерную плотность (2.41).

Для примера укажем характеристики гауссовского белого шума, прошедшего через полосовой фильтр с граничными частотами ω_1, ω_2, модулем коэффициента передачи $\mathscr{K}_0$ и полосой пропускания $\Delta\omega=\omega_2-\omega_1$. Полосовым фильтром белый шум преобразуется в узкополосную помеху — гауссовский стационарный узкополосный процесс. Дисперсия помехи на выходе фильтра

$$\mathscr{D}_2 = 2\Omega\mathscr{K}^2{}_0\Delta f = \Omega_0\mathscr{K}^2{}_0\Delta f = \mathscr{K}^2{}_0\mathscr{D}_1, \tag{4.43}$$

где Ω_0 — спектральная плотность белого шума в области положительных частот; $\Delta f=\Delta\omega/2\pi$. Корреляционная функция помехи определяется по формуле (2.74), где $\mathscr{D}=\mathscr{D}_2$.

Таким образом, анализ прохождения случайных сигналов через непрерывные каналы связи по существу сводится к последовательному анализу методами статистической радиотехники и теории радиотехнических цепей различных функциональных преобразований случайных сигналов и последовательностей. Результаты анализа используют и для определения характеристик дискретных каналов.

Контрольные вопросы

1. Сформулируйте постановку задачи анализа прохождения случайных сигналов через каналы.
2. В чем особенности анализа характеристик сигналов на выходе линейных и нелинейных систем?

3. Какой метод обычно используют для анализа прохождения случайных сигналов через линейные инерционные системы?
4. В чем сущность эффекта нормализации случайных сигналов в линейных узкополосных системах?
5. Поясните, как изменяются характеристики белого шума после прохождения полосового фильтра.
6. Как определяют одномерную плотность распределения сигнала на выходе нелинейной безынерционной системы?
7. Как определяют корреляционную функцию и спектральную плотность сигнала на выходе нелинейной безынерционной системы?

4.5. ВЫВОДЫ

1. Целью анализа непрерывных каналов является анализ искажений сигналов и влияния помех в каналах на передачу сигналов. Реальные непрерывные каналы являются нелинейными инерционными стохастическими системами. Анализ таких систем при случайных входных воздействиях представляет сложную проблему, которая является предметом современных научных исследований. Для приближенного решения задач анализа непрерывные каналы представляют в виде последовательного соединения линейной инерционной системы и нелинейной безынерционной системы. Это позволяет для многих практически важных случаев оценить линейные и нелинейные искажения сигналов в каналах и влияние помех. Каналы рассматривают как динамические системы, реакция на выходе которых не может предшествовать воздействию на входе.

Помехи в непрерывных каналах бывают мультипликативные и аддитивные. Мультипликативные помехи обусловлены случайными изменениями коэффициента передачи канала из-за изменения характеристик линий связи. Аддитивные помехи вызваны флуктуационными явлениями, связанными с тепловыми процессами в элементах аппаратуры, атмосферными и индустриальными процессами, работой соседних каналов. Математическими моделями помех служат узкополосные случайные сигналы, случайные последовательности импульсов, гауссовский белый шум и др.

Для анализа каналов разрабатывают математические модели каналов, которые включают описание помех, структур и параметров линейных и нелинейных операторов, отражающих преобразования сигналов в каналах. Математические модели отличаются сложностью математического описания работы канала, требуемыми исходными данными для анализа, погрешностями описания и получаемыми с помощью моделей результатами. Принимают следующие модели: идеальный канал, гауссов канал без искажений сигналов, гауссов канал с неопределенной фазой, гауссов однолучевой канал с замираниями сигналов, гауссов многолучевой канал с замираниями и сосредоточенными аддитивными помехами.

Для анализа конкретных каналов обычно выбирают такие модели, которые приводят к относительно простым решениям задач анализа, но обеспечивают точность, требуемую для инженерных расчетов. Наибольшее распространение получили различные раз-

новидности гауссова канала. В этих моделях предполагается, что могут быть линейные и нелинейные искажения сигналов в каналах, аддитивная помеха является гауссовым стационарным случайным процессом.

2. Основными характеристиками дискретно-непрерывного канала являются алфавит и вероятности появления входных кодовых символов, полоса пропускания непрерывного канала, на котором построен дискретно-непрерывный канал, априорная плотность $f(s|b_{k1})$ распределения вероятности появления сигнала $s_2(t)$ при условии, что передавался сигнал b_{k1}.

Результатом анализа дискретно-непрерывного канала является определение апостериорной вероятности $P(b_{k1}|s)$ того, что при полученном сигнале $s_2(t)$ передавался кодовый символ b_{k1}. Эту вероятность определяют по формуле Байеса (4.14). Если решающая схема демодулятора работает по алгоритму определения максимума апостериорной вероятности, на выходе демодулятора появляется тот символ, апостериорная вероятность появления которого больше других.

Дискретно-непрерывный канал называют стационарным, если условная плотность $f(s|b_{k1})$ не зависит от времени. Если она не зависит от того, какие символы передавались до b_{k1}, то такой канал называют каналом без памяти. Если дискретно-непрерывный канал построен на гауссовском канале, он является стационарным и без памяти. Стационарные дискретно-непрерывные каналы без памяти получили широкое применение для приближенного анализа реальных каналов.

3. Основными характеристиками дискретного канала являются алфавит и априорные вероятности появления входных кодовых символов, скорость передачи этих символов, алфавит символов копии сообщений, условная вероятность $P(a_{i2}|a_{k1})$ появления символа a_{i2} при условии, что был передан a_{k1}. Результатом анализа дискретного канала является определение апостериорной вероятности $P(a_{k1}|a_{i2})$ того, что при полученном символе a_{i2} передавался символ a_{k1}. Апостериорные вероятности рассчитывают по формуле Байеса (4.19). Если решающая схема декодера работает по алгоритму определения максимума апостериорной вероятности, на выходе декодера появляется тот символ, вероятность появления которого больше других.

Дискретный канал называют однородным, если для любых сочетаний a_{i2} и a_{k1} условная вероятность $P(a_{i2}|a_{k1})$ не зависит от момента взятия отсчета. Если она не зависит от того, какие символы передавались до a_{k1}, то такой канал называют каналом без памяти. Если дискретный канал построен на гауссовском канале, он является однородным и без памяти. Однородные дискретные каналы без памяти получили широкое распространение для приближенного анализа реальных каналов.

Апостериорные вероятности $P(a_{k1}|a_{i2})$ и априорные вероятности $P(a_{k1})$ используют для расчета полной вероятности появления ошибки в канале, полной вероятности правильного приема, веро-

ятностей появления различных символов на выходе канала в установившемся режиме, скорости передачи информации, пропускной способности канала и других характеристик. Математическими методами определения характеристик дискретных каналов являются методы теории марковских цепей [9, 15, 18].

В связи с широким применением двоичных кодов важное значение имеют двоичные дискретные каналы. Наиболее распространенной является модель однородного двоичного канала без памяти. Если условные вероятности появления ошибок равны $p_{12}=p_{21}=p_0$, канал называют симметричным. Если для передачи используют противоположные сигналы с одинаковыми амплитудами, то условная вероятность появления ошибки в канале полностью определяется отношением амплитуды сигнала к среднеквадратическому значению аддитивной помехи в гауссовском канале. В силу симметрии канала полная (безусловная) вероятность появления ошибки совпадает с условной вероятностью появления ошибки.

При $p_0 \ll 1$ и передаче независимых кодовых сигналов вероятность появления нескольких ошибок в одной кодовой комбинации падает с ростом числа ошибок. Максимальной является вероятность того, что ошибок не будет. Поэтому основное внимание уделяют обнаружению и исправлению одиночных ошибок и других ошибок малой кратности. При коррелированных кодовых сигналах (в каналах с памятью) наблюдаются иные распределения вероятности появления кратных ошибок.

4. Прохождение сигналов через каналы обычно изучают методами статистической радиотехники [15] и теории радиотехнических цепей [7]. Если канал рассматривают как последовательное соединение линейной инерционной системы и нелинейной безынерционной системы, то используют известные в статистической радиотехнике результаты анализа прохождения случайных сигналов через линейные и нелинейные устройства.

Задача анализа прохождения сигналов через канал в полной мере решена тогда, когда по заданной n-мерной плотности распределения входного сигнала и оператору преобразования сигнала в канале найдена n-мерная плотность распределения выходного сигнала. На практике используют и определяют более простые характеристики сигналов: одномерные и двумерные плотности распределения, корреляционные функции и спектральные плотности.

В линейных системах корреляционные функции и спектральные плотности выходных сигналов определить относительно просто, сложнее находить плотности распределения выходных сигналов. Положение облегчают два обстоятельства. Если входной сигнал является гауссовским, то выходной — также гауссов. Если ширина спектра входного сигнала значительно превышает полосу пропускания системы, то вне зависимости от закона распределения входного процесса плотность распределения выходного будет близка к гауссовской. Этот эффект называют нормализацией случайных процессов в линейных системах.

Обычно для выходных сигналов определяют методом Дюамеля корреляционную функцию и по ней с помощью соотношения Хинчина—Винера спектральную плотность. В линейных каналах прохождение сигнала и аддитивной помехи рассматривают отдельно и затем используют метод суперпозиции.

Белый шум, прошедший через полосовой фильтр, является узкополосным гауссовским случайным процессом со средней частотой спектра, равной средней частоте полосы пропускания фильтра. Дисперсия выходного сигнала (4.43) равна дисперсии белого шума, отсчитанной по полосе пропускания фильтра и умноженной на квадрат коэффициента передачи фильтра. Дисперсия прямо пропорциональна спектральной плотности белого шума и полосе пропускания фильтра.

В нелинейных безынерционных системах в отличие от линейных плотности распределения выходных сигналов определить относительно просто, сложнее находить корреляционные функции и спектральные плотности. После прохождения гауссовского случайного сигнала через нелинейный канал его распределение отличается от гауссовского. Поэтому в данном случае нельзя найти плотность распределения выходного сигнала по его корреляционной функции. Прохождение сигнала и аддитивной помехи через нелинейный канал уже нельзя рассматривать отдельно, так как метод суперпозиции несправедлив. Обычно отыскивают корреляционную функцию выходного сигнала, а потом с помощью соотношения Хинчина—Винера определяют его спектральную плотность.

Глава 5

ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ИСТОЧНИКОВ СООБЩЕНИЙ И КАНАЛОВ

5.1. ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ИСТОЧНИКОВ ДИСКРЕТНЫХ СООБЩЕНИЙ

Информационные характеристики источников дискретных сообщений были определены в § 1.4. Их анализ показывает, что условная энтропия источника дискретных сообщений является его основной информационной характеристикой, через которую выражают большинство других. Поэтому целесообразно рассмотреть свойства условной энтропии и показать, как влияют на нее неравномерность распределения вероятностей появления символов сообщений и статистические связи между символами.

Если появление символа a_i зависит только от того, какой был в сообщении предыдущий символ a_j, то образование сообщений описывают простой марковской цепью (см. п. 4.3.2). Энтропию совместного появления двух символов определяют, применяя операцию усреднения по всему объему алфавита:

$$H(\mathrm{A}, \mathrm{A}') = -\sum_{i=1}^{m}\sum_{j=1}^{m} p(a_i, a'_j)\log p(a_i, a'_j), \qquad (5.1)$$

где $p(a_i, a'_j)$ — вероятность совместного появления символов a_i и a'_j; $\log p(a_i, a'_j)$ — количество информации, которое приходится на слог $a_i a'_j$.

Так как

$$p(a_i, a'_j) = p(a_i)\,p(a'_j|a_i) = p(a'_j)\,p(a_i|a'_j), \qquad (5.2)$$

где $p(a_i)$, $p(a'_j)$ — вероятности появления a_i, a'_j; $p(a'_j|a_i)$ — вероятность появления a'_j при условии, что перед ним появился a_i; $p(a_i|a'_j)$ — вероятность появления a_i при условии, что перед ним появился a'_j, то выражение (5.1) можно представить в виде

$$H(\mathrm{A}, \mathrm{A}') = -\sum_{i=1}^{m}\sum_{j=1}^{m} p(a_i)\,p(a'_j|a_i)\log p(a_i)\,p(a'_j|a_i) =$$

$$= -\sum_{i=1}^{m} p(a_i)\log p(a_i)\sum_{j=1}^{m} p(a'_j|a_i) - \sum_{i=1}^{m} p(a_i)\sum_{j=1}^{m} p(a'_j|a_i)\log p(a'_j|a_i).$$

В соответствии с условием нормировки (4.25) $\sum_{j=1}^{m} p(a'_j | a_i) = 1$, поэтому

$$H(\mathbf{A},\ \mathbf{A}') = H(\mathbf{A}) + H(\mathbf{A}' | \mathbf{A}), \tag{5.3}$$

где $H(\mathbf{A})$ — энтропия источника (1.8); $H(\mathbf{A}'|\mathbf{A})$ — условная энтропия источника. Используя второе произведение в (5.2), аналогично получим

$$H(\mathbf{A},\ \mathbf{A}') = H(\mathbf{A}') + H(\mathbf{A} | \mathbf{A}'). \tag{5.4}$$

Следовательно, среднее количество информации, которое приносят два соседних символа, равно сумме среднего количества информации, которую приносит первый из них, и среднего количества информации, которую приносит второй, при условии, что первый уже появился.

Поэтому условная энтропия одного символа (1.10) — это то среднее количество информации, которое приносит последующий символ при условии, что предыдущий уже известен:

$$H(\mathbf{A} | \mathbf{A}') = H(\mathbf{A},\ \mathbf{A}') - H(\mathbf{A}'), \tag{5.5}$$

$$H(\mathbf{A}' | \mathbf{A}) = H(\mathbf{A},\ \mathbf{A}') - H(\mathbf{A}). \tag{5.6}$$

Условная энтропия обладает следующим свойством: если символы a_i и a'_j взаимозависимы, то

$$H(\mathbf{A} | \mathbf{A}') < H(\mathbf{A}). \tag{5.7}$$

Следовательно,

$$H(\mathbf{A}') + H(\mathbf{A} | \mathbf{A}') < H(\mathbf{A}') + H(\mathbf{A}). \tag{5.8}$$

Только для источников с независимыми символами

$$H(\mathbf{A} | \mathbf{A}') = H(\mathbf{A}), \tag{5.9}$$

$$H(\mathbf{A},\ \mathbf{A}') = H(\mathbf{A}) + H(\mathbf{A}'). \tag{5.10}$$

Так как

$$\max_{p(a_j)} H(\mathbf{A}) = \log m, \tag{5.11}$$

то существуют две причины, из-за которых уменьшается среднее количество информации, переносимое одним символом в реальных сообщениях: неравномерность распределения вероятностей появления различных символов и корреляционные связи между символами. В реальных сообщениях корреляционные связи существуют не только между двумя соседними символами, а и между $l+1$ символами. Говорят, что источник имеет память на l символов. Формула (1.10) выведена для случая $l=1$. Аналогично определяют условную энтропию для более общих случаев, когда $l>1$. Для реальных источников энтропия (1.8) и производитель-

ность (1.12) отличаются от максимальных, а избыточность (1.11) велика.

Рассмотрим *двоичный источник* и покажем, как влияют на энтропию, производительность и избыточность неравновероятное появление символов и корреляционные связи между ними. Обозначим символы источника через a_1, a_2, вероятности их появления через $p(a_1)$, $p(a_2)$, соответствующие условные вероятности через $p(a_1|a'_1)$, $p(a_2|a'_2)$, $p(a_1|a'_2)$, $p(a_2|a'_1)$.

5.1.1. Двоичный источник с независимыми равновероятными символами. В этом случае $p(a_1)=p(a_2)=p=1/2$, условные вероятности равны нулю. Энтропия такого источника максимальна:

$$H_{\text{макс}}=-\frac{1}{2}\log\frac{1}{2}-\frac{1}{2}\log\frac{1}{2}=-\log\frac{1}{2}=\log 2=1 \text{ бит/симв.} \quad (5.12)$$

Следовательно, *1 бит* — это максимальное среднее количество информации, которое может переносить один символ источника двоичных сообщений. Производительность такого источника максимальна, избыточность отсутствует.

5.1.2. Двоичный источник с независимыми неравновероятными символами. В этом случае $p(a_1)=p$, $p(a_2)=1-p$, условные вероятности равны нулю. Энтропия такого источника

$$H(p)=-p\log p-(1-p)\log(1-p). \quad (5.13)$$

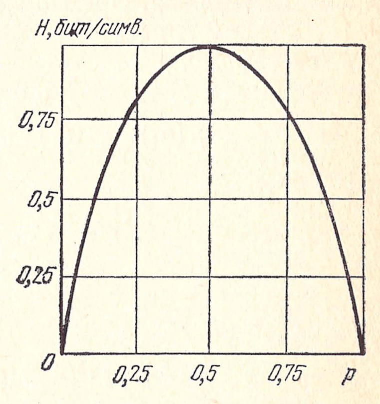


Рис. 5.1. График функции (5.13)

На рис. 5.1 показан график функции $H(p)$. Эта функция достигает максимума $H_{\text{макс}}=1$ бит/симв. при $p=1/2$ и равна нулю при $p=0$ и $p=1$. Так как $H(p)<H_{\text{макс}}$ при $p\neq 1/2$, то производительность такого источника меньше максимальной, а избыточность

$$r(p)=1-H(p)/H_{\text{макс}} \quad (5.14)$$

больше нуля.

Например, если $p(a_1)=0{,}125$, $p(a_2)=0{,}875$, то $H(p)=-0{,}125\log 0{,}125-0{,}875\log 0{,}875\simeq 0{,}576$ бит/симв.,

$$r(p)=1-0{,}576\simeq 0{,}42.$$

Следовательно, отклонение распределения вероятностей появления символов от равномерного приводит к снижению энтропии и увеличению избыточности источника.

5.1.3. Двоичный источник с коррелированными равновероятными символами. В этом случае $p(a_1)=p(a_2)=p=1/2$, условные вероятности не равны нулю. Предположим, что $p(a_1|a'_1)=p(a_2|a'_2)$, а $p(a_1|a'_2)=p(a_2|a'_1)$. Условную энтропию источника

найдем по формуле (1.10):

$$H(\mathbf{A}|\mathbf{A}') = -p[p(a_1|a'_1)\log p(a_1|a'_1) + p(a_2|a'_1)\log p(a_2|a'_1)] - p[p(a_2|a'_2)\log p(a_2|a'_2) + p(a_1|a'_2)\log p(a_1|a'_2)] = 2p[p(a_1|a'_1)\log p(a_1|a'_1) + p(a_2|a'_1)\log p(a_2|a'_1)]. \quad (5.15)$$

Пусть $p(a_1|a'_1)=0{,}7$, $p(a_2|a'_1)=0{,}3$, тогда

$$H(\mathbf{A}|\mathbf{A}') = -2\frac{1}{2}[0{,}7\log 0{,}7 + 0{,}3\log 0{,}3] \approx 0{,}883 \text{ бит/симв.},$$

$$r = 1 - 0{,}883 \approx 0{,}12.$$

Следовательно, наличие статистических связей между символами также приводит к уменьшению энтропии и увеличению избыточности источника.

5.1.4. Двоичный источник с коррелированными неравновероятными символами. Предположим, что $p(a_1|a'_2)=0{,}1$, $p(a_1|a'_1)=0{,}3$, $p(a_2|a'_1)=0{,}7$, $p(a_2|a'_2)=0{,}9$. Для определения условной энтропии необходимо определить вначале вероятности $p(a_1)$ и $p(a_2)$. Воспользуемся формулой полной вероятности, тогда

$$p(a_1) = p(a_1)p(a_1|a'_1) + [1 - p(a_1)]p(a_1|a'_2).$$

Отсюда

$$p(a_1) = \frac{p(a_1|a'_2)}{1 + p(a_1|a'_2) - p(a_1|a'_1)},$$

$$p(a_2) = 1 - p(a_1).$$

Условная энтропия

$$H(\mathbf{A}|\mathbf{A}') = -p(a_1)[p(a_1|a'_1)\log p(a_1|a'_1) + p(a_2|a'_1)\log p(a_2|a'_1)] - p(a_2)[p(a_2|a'_2)\log p(a_2|a'_2) + p(a_1|a'_2)\log p(a_1|a'_2)]. \quad (5.16)$$

Подставив численные значения вероятностей, получим

$$H(\mathbf{A}|\mathbf{A}') = -0{,}125(0{,}3\log 0{,}3 + 0{,}7\log 0{,}7) - 0{,}875(0{,}1\log 0{,}1 + 0{,}9\log 0{,}9) \simeq 0{,}51 \text{ бит./симв.}$$

Сравнение с данными п. 5.1.2 показывает, что корреляция соседних символов источника приводит к снижению энтропии и увеличению избыточности. Из-за влияния корреляции и неравномерности распределения вероятностей появления символов энтропия в приведенном примере примерно вдвое меньше максимальной, а избыточность примерно 49% (см. п. 5.1.1).

Рассмотрение информационных характеристик двоичного источника показывает, что увеличить энтропию источника и уменьшить избыточность его сообщений можно двумя способами: «выравниванием» распределения вероятностей появления символов

и «декорреляцией» символов сообщений — устранением корреляционных связей между символами. Эти идеи лежат в основе построения оптимальных эффективных кодов (см. § 5.3).

Контрольные вопросы

1. Как определяют энтропию двух символов?
2. Как определяют условную энтропию источника с памятью на один символ?
3. В чем причины того, что условная энтропия реальных источников отличается от максимальной?
4. Как рассчитать условную энтропию, производительность и избыточность двоичного источника с неравновероятными и коррелированными символами?
5. Как повысить производительность источников дискретных сообщений?

5.2. ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ДИСКРЕТНЫХ КАНАЛОВ

Информационные характеристики дискретных каналов определены в § 1.4. Кроме того, применяют еще одну характеристику — коэффициент использования канала

$$\eta = R/C. \tag{5.17}$$

Так как $0 \leqslant R \leqslant C$, то $0 \leqslant \eta \leqslant 1$. Коэффициент использования канала показывает, в какой степени скорость передачи информации приближается к пропускной способности канала.

Рассмотрим информационные характеристики дискретных каналов без ошибок и с ошибками, определим избыточность кодов и длину кодовых комбинаций в дискретных каналах с ошибками, а также пропускную способность двоичного и m-ичного каналов с ошибками.

5.2.1. Идеальные дискретные каналы. Идеальным называют канал, ошибки в котором отсутствуют. Для передачи сообщений по каналу символы кодируют и преобразуют в электрические кодовые сигналы. Эту операцию выполняет кодер (см. рис. 1.1). В идеальном канале между элементами кодовых сигналов на входе и элементами сигналов на выходе существует однозначное соответствие. Скорость передачи информации равна производительности кодера

$$R = H'_{\text{к}} = W_{\text{к}} H_{\text{к}} \quad [\text{бит/с}], \tag{5.18}$$

где $W_{\text{к}} = 1/\tau$ — скорость передачи кодовых сигналов (сигн./с); $H_{\text{к}}$ — энтропия кодера (бит/сигн.); τ — длительность элементарного кодового сигнала.

В соответствии с формулой (1.14) пропускная способность идеального канала

$$C = \max R = W_{\text{к}} \log m_{\text{к}}, \tag{5.19}$$

где $m_{\text{к}}$ — основание кода. Пропускная способность является предельной характеристикой канала. Покажем это. Если основание кода равно $m_{\text{к}}$, а для передачи одного элементарного кодового сигнала необходимо время τ, то для передачи кодовой комбинации

длиной n элементарных сигналов потребуется время $T=n\tau$. Общее число кодовых комбинаций длительностью T, как обычно, равно $N(T)=m^{n}_{\text{к}}$. Следовательно, максимальное количество информации в одной кодовой комбинации $H_{\text{макс}}=n\log m_{\text{к}}$. Пропускная способность

$$C=\lim_{n\to\infty}\frac{n\log m_{\text{к}}}{n\tau}=W_{\text{к}}\log m_{\text{к}}.$$

Таким образом, пропускную способность идеального дискретного канала полностью определяют скорость передачи сигналов и основание кода.

К. Шеннон доказал следующую *теорему*: если ошибки в дискретном канале отсутствуют, можно закодировать сообщение на выходе источника так, чтобы передавать информацию со средней скоростью R, сколь угодно близкой к C. Передавать информацию с $R>C$ невозможно.

Эта теорема служит теоретической основой для построения оптимальных эффективных кодов (см. § 5.3). Если в процессе кодирования на выходе кодера обеспечить появление равновероятных независимых кодовых сигналов, то каждый элементарный сигнал будет нести максимальное количество информации, производительность кодера будет максимальной и скорость передачи информации приблизится к пропускной способности.

5.2.2. Реальные дискретные каналы. В реальных каналах всегда имеются ошибки и стирания символов (см. § 4.3) при передаче сообщений; вероятности появления ошибок во многом определяются искажениями сигналов и помехами в непрерывных каналах, на основе которых построены дискретные каналы, и рядом других причин. В реальных каналах передаваемые сигналы $s_1(t)$ (см. рис. 1.1 и § 4.4) искажаются и на демодулятор $D_{\text{м2}}$ поступают принятые сигналы $s_2(t)$, которые в той или иной мере отличаются от передаваемых. На выходе решающей схемы может появиться кодовый сигнал b_{k2}, который будет отличаться от переданного b_{i1}. В этом случае говорят, что произошла ошибка. Ошибки приводят к уменьшению пропускной способности каналов и потере информации.

Количество информации, которое содержит принятый символ относительно переданного или в более общем случае один символ относительно другого, находят с помощью формулы (5.2) для вероятности совместного появления символов. Когда символы появляются независимо, условные вероятности в (5.2) являются безусловными и вероятность совместного появления символов определяется как произведение вероятностей появления каждого символа. В этом единственном случае один символ не несет никакой информации о другом.

Поэтому количество информации, которое содержит принятый сигнал b_{k2} относительно переданного b_{i1}, с учетом (5.2) опре-

деляют по формуле

$$I(b_{k2}, b_{i1}) = \log p(b_{k2}, b_{i1}) - \log[p(b_{k2})\, p(b_{i1})] =$$
$$= \log \frac{p(b_{k2}, b_{i1})}{p(b_{i1})\, p(b_{k2})} = \log \frac{p(b_{i1} \mid b_{k2})}{p(b_{i1})} = \log \frac{p(b_{k2} \mid b_{i1})}{p(b_{k2})}, \qquad (5.20)$$

где $p(b_{k2}, b_{i1})$ — вероятность совместного появления b_{i1}, b_{k2}; $p(b_{i1})$, $p(b_{k2})$ — вероятности появления b_{i1}, b_{k2}; $p(b_{i1}|b_{k2})$, $p(b_{k2}|b_{i1})$ — соответствующие условные вероятности. Если символы появляются независимо, то $I(b_{k2}, b_{i1}) = \log 1 = 0$. Во всех остальных случаях один символ несет информацию о другом и $I(b_{k2}, b_{i1}) \neq 0$.

Среднее количество принятой информации, которое приносит один символ, получим, усредняя (5.20) по всем i и k,

$$I(\mathrm{B}_2, \mathrm{B}_1) = \sum_{i=1}^{m} \sum_{k=1}^{m} p(b_{k2}, b_{i1}) \log \frac{p(b_{k2}, b_{i1})}{p(b_{i1})\, p(b_{k2})}. \qquad (5.21)$$

Учтя две формы записи дроби (5.20), получим две формы записи для среднего количества информации в принятом символе

$$I(\mathrm{B}_2, \mathrm{B}_1) = -\sum_{i=1}^{m} p(b_{i1}) \log p(b_{i1}) + \sum_{k=1}^{m} p(b_{k2}) \sum_{i=1}^{m} p(b_{i1} \mid b_{k2}) \times$$
$$\times \log p(b_{i1} \mid b_{k2}), \qquad (5.22)$$

$$I(\mathrm{B}_2, \mathrm{B}_1) = -\sum_{k=1}^{m} p(b_{k2}) \log p(b_{k2}) + \sum_{i=1}^{m} p(b_{i1}) \times$$
$$\times \sum_{k=1}^{m} p(b_{k2} \mid b_{i1}) \log p(b_{k2} \mid b_{i1}). \qquad (5.23)$$

Выражения (5.22), (5.23) можно записать более наглядно

$$I(\mathrm{B}_2, \mathrm{B}_1) = H(\mathrm{B}_1) - H(\mathrm{B}_1 \mid \mathrm{B}_2), \qquad (5.24)$$

$$I(\mathrm{B}_2, \mathrm{B}_1) = H(\mathrm{B}_2) - H(\mathrm{B}_2 \mid \mathrm{B}_1). \qquad (5.25)$$

Смысл выражений (5.24), (5.25) следующий. Величина $H(\mathrm{B}_1)$ — это энтропия кодера, а величина $H(\mathrm{B}_1 | \mathrm{B}_2)$ — это среднее количество информации, потерянное в канале из-за ошибок. Следовательно, соотношение (5.24) показывает, что среднее количество принятой в одном символе информации равно разности среднего количества переданной информации и среднего количества информации, потерянной в канале из-за ошибок. Соотношение (5.25) показывает, что среднее количество принятой в одном символе информации можно вычислить и как разность энтропий принятого сигнала и помехи. Соотношение (5.25) используют чаще, так как оно позволяет определить $I(\mathrm{B}_2, \mathrm{B}_1)$ через энтропию помехи, которую определить проще.

В соответствии с (1.13) *скорость передачи информации* в реальных каналах $R=W_к I(\mathbf{B}_2, \mathbf{B}_1)$. Использовав (5.24) и (5.25), получим

$$R=W_к[H(\mathbf{B}_1)-H(\mathbf{B}_1|\mathbf{B}_2)]=W_к[H(\mathbf{B}_2)-H(\mathbf{B}_2|\mathbf{B}_1)]. \quad (5.26)$$

Если ошибок нет, то $H(\mathbf{B}_1|\mathbf{B}_2)=H(\mathbf{B}_2|\mathbf{B}_1)=0$ и формула (5.26) переходит в формулу (5.18) для идеального канала.

Пропускная способность реальных дискретных каналов

$$C=\max R=\max W_к[H(\mathbf{B}_1)-H(\mathbf{B}_1|B_2)]= \\ =\max W_к\,[H(\mathbf{B}_2)-H(\mathbf{B}_2|\mathbf{B}_1)], \quad (5.27)$$

где операция отыскания максимума выполняется по всем способам передачи и обработки сигналов.

Для реальных дискретных каналов К. Шеннон доказал следующую *теорему*: если производительность источника сообщений меньше пропускной способности канала, сообщение можно закодировать в сигналы так, чтобы передавать информацию по дискретному каналу с помехами со сколь угодно малой вероятностью ошибки. Эта теорема является теоретической основой корректирующего кодирования. В ней утверждается, что существует такой код, использование которого позволит обнаружить и исправить практически все ошибки. Задача заключается в отыскании и построении таких кодов.

5.2.3. Избыточность кодов и длина кодовых комбинаций в реальных каналах. Установим связь, которая должна существовать в реальных каналах для обеспечения сколь угодно высокой верности между средней длиной n_0 кодовой комбинации, избыточностью кода $r_к$ и количеством $H(\mathbf{B}_1|\mathbf{B}_2)$ информации, теряемой из-за помех.

Чтобы кодер успевал преобразовать каждый символ сообщения в кодовую комбинацию, средней длиной n_0 элементарных кодовых сигналов, скорость $W_к$ передачи сигналов кодером должна быть в n_0 раз выше скорости $W_и$ передачи символов источником. Поэтому для безошибочного кодирования должно выполняться условие

$$W_к=n_0 W_и. \quad (5.28)$$

Кроме этого условия, должно выполняться и условие отсутствия потерь информации при кодировании

$$H_и(\mathbf{B}_1)=n_0 H(\mathbf{B}_1). \quad (5.29)$$

Это условие о том, что среднее количество информации $H_п(\mathbf{B}_1)$, которое заключено в одном символе сообщения, должны переносить n_0 символов кодовой комбинации. Учтем (5.29), тогда избыточность кода для реальных каналов

$$r_к=1-H(\mathbf{B}_1)/\log m_к=1-H_и(\mathbf{B}_1)/n_0 \log m_к. \quad (5.30)$$

Условие теоремы Шеннона для реальных каналов с учетом (5.27) можно представить в виде

$$W_и H_и(\mathbf{B}_1) < W_к \max[H(\mathbf{B}_1) - H(\mathbf{B}_1|\mathbf{B}_2)]$$

или иначе

$$\frac{W_и H_и(\mathbf{B}_1)}{W_к \max[H(\mathbf{B}_1) - H(\mathbf{B}_1|\mathbf{B}_2)]} < 1. \quad (5.31)$$

Учтя (5.28) и разрешив неравенство (5.31) относительно n_0, получим

$$n_0 > \frac{H_и(\mathbf{B}_1)}{\max[H(\mathbf{B}_1) - H(\mathbf{B}_1|\mathbf{B}_2)]}. \quad (5.32)$$

Из неравенства (5.32) следует практически важный вывод: с ростом среднего количества информации $H(\mathbf{B}_1|\mathbf{B}_2)$, теряемой в канале из-за помех, для обеспечения сколь угодно высокой верности передачи информации должна расти средняя длина кодовой комбинации.

Аналогичный вывод справедлив и относительно избыточности кода (5.30). Если n_0 растет, дробь в правой части (5.30) уменьшается, а значение $r_к$ увеличивается. Можно установить и непосредственную связь между $r_к$ и $H(\mathbf{B}_1|\mathbf{B}_2)$. Так как $\max[H(\mathbf{B}_1) - H(\mathbf{B}_1|\mathbf{B}_2)] = \log m_к - H(\mathbf{B}_1|\mathbf{B}_2)$, то неравенство (5.31) можно представить в виде

$$W_и H_и(\mathbf{B}_1) < W_к \log m_к - W_к H(\mathbf{B}_1|\mathbf{B}_2). \quad (5.33)$$

Разделив обе части неравенства (5.33) на $W_к \log m_к$, получим

$$\frac{W_и H_и(\mathbf{B}_1)}{W_к \log m_к} < 1 - \frac{H(\mathbf{B}_1|\mathbf{B}_2)}{\log m_к}. \quad (5.34)$$

Учтя (5.28) и поменяв местами дроби в неравенстве, получим

$$1 - H_и(\mathbf{B}_1)/n_0 \log m_к > H(\mathbf{B}_1|\mathbf{B}_2)/\log m_к.$$

Левая часть неравенства — коэффициент избыточности кода (5.30). Следовательно, для обеспечения сколь угодно высокой верности передачи информации в реальных каналах должно выполняться неравенство

$$r_к > H(\mathbf{B}_1|\mathbf{B}_2)/\log m_к. \quad (5.35)$$

Таким образом, для обеспечения сколь угодно высокой верности передачи информации в реальных каналах с ростом потерь информации $H(\mathbf{B}_1|\mathbf{B}_2)$ из-за помех должны расти средняя длина кодовой комбинации и избыточность кода.

5.2.4. Пропускная способность двоичного и m-ичного реальных каналов. Определим с помощью соотношения (5.27) пропускную способность реального двоичного симметричного канала без памяти. Предположим, что известна вероятность p_0 появления ошибки в канале (см. п. 4.3.3).

Определим значение $\max[H(\mathbf{B}_2) - H(\mathbf{B}_2|\mathbf{B}_1)]$. Как следует из (5.12), $\max H(\mathbf{B}_2) = \log 2 = 1$ бит/сигн. Условная энтропия

$H(\mathrm{B}_2|\mathrm{B}_1)$ — это энтропия помехи [1—3], которая определяется по формуле (5.16):

$$H(\mathrm{B}_2|\mathrm{B}_1) = -p(b_1)[p(b_1|b_1)\log p(b_1|b_1) + p(b_2|b_1)\log p(b_2|b_1)] - p(b_2)[p(b_2|b_2)\log p(b_2|b_2) + p(b_1|b_2)\log p(b_1|b_2)]. \quad (5.36)$$

Подставив значения условных вероятностей появления ошибок, получим

$$H(\mathrm{B}_2|\mathrm{B}_1) = -p(b_1)[(1-p_0)\log(1-p_0) + p_0\log p_0] - p(b_2)[(1-p_0)\log(1-p_0) + p_0\log p_0] = -[p(b_1) + p(b_2)][(1-p_0)\log(1-p_0) + p_0\log p_0].$$

Так как по условию нормировки (4.25) первая сумма равна единице, то

$$H(\mathrm{B}_2|\mathrm{B}_1) = -[p_0\log p_0 + (1-p_0)\log(1-p_0)]. \quad (5.37)$$

Пропускная способность двоичного реального канала

$$C_2 = W_\text{к}[1 + p_0\log p_0 + (1-p_0)\log(1-p_0)]. \quad (5.38)$$

Анализ зависимости $C_2(p_0)$ показывает, что в диапазоне реальных изменений $p_0 \in [0;\ 0{,}5]$ функция $C_2(p_0)$ является монотонно убывающей. При $p_0 = 0{,}5$, $C_2(p_0) = 0$ это означает, что из-за высокого уровня помех в канале кодовые сигналы на входе и на выходе канала становятся независимыми (принимаемые сигналы не несут информации о передаваемых).

Пропускную способность m-ичного реального канала определяют аналогично

$$C_m = W_\text{к}[\log m_\text{к} + p_0\log p_0(m_\text{к}-1)^{-1} + (1-p_0)\log(1-p_0)]. \quad (5.39)$$

Из (5.39) как частный случай следует (5.38) при $m_\text{к}=2$. Если $p_0 \to 0$, то пропускная способность реального канала стремится к пропускной способности идеального канала (5.19).

Средняя длина кодовых комбинаций в двоичном и m-ичном реальных каналах определяется неравенством (5.32):

$$n_{\text{о}2} > \frac{H_\text{и}(\mathrm{B}_1)}{1 + p_0\log p_0 + (1-p_0)\log(1-p_0)},$$

$$n_{\text{о}m} > \frac{H_\text{и}(\mathrm{B}_1)}{\log m_\text{к} + p_0\log p_0(m_\text{к}-1)^{-1} + (1-p_0)\log(1-p_0)}. \quad (5.40)$$

Следовательно, минимальная средняя длина кодовых комбинаций в реальных каналах определяется энтропией источника, основанием кода и вероятностью появления ошибки в канале при передаче одного кодового сигнала.

Избыточность двоичного кода (см. (5.35))

$$r_2 > p_0\log p_0 + (1-p_0)\log(1-p_0), \quad (5.41)$$

избыточность многопозиционного кода

$$r_m > \frac{p_0 \log p_0 (m_{\text{к}} - 1)^{-1} + (1 - p_0) \log (1 - p_0)}{\log m_{\text{к}}}.$$

Контрольные вопросы

1. Как определяют скорость передачи информации и пропускную способность идеального дискретного канала?
2. Сформулируйте и поясните теорему кодирования Шеннона для идеального дискретного канала.
3. Как определить количество информации в принятом сигнале относительно переданного по реальному каналу?
4. Как определить для реального канала среднее количество информации в одном принятом кодовом сигнале?
5. Напишите и поясните соотношения для скорости передачи информации и пропускной способности реального канала.
6. Сформулируйте и поясните теорему кодирования Шеннона для реального дискретного канала.
7. Напишите и поясните формулы для избыточности кода и средней минимальной длины кодовых комбинаций для реальных дискретных каналов.
8. Как определяют пропускную способность двоичного и m-ичного реальных дискретных каналов?
9. От каких характеристик зависит средняя минимальная длина кодовых комбинаций в реальном m-ичном канале?

5.3. ОПТИМАЛЬНОЕ ЭФФЕКТИВНОЕ КОДИРОВАНИЕ

Оптимальное эффективное кодирование позволяет согласовать источник с каналом и обеспечить наилучшее использование пропускной способности канала. Сущность эффективного кодирования заключается в том, что неравномерное распределение вероятностей появления коррелированных символов сообщений с помощью определенным образом выбранного кода переводят в равномерное распределение вероятностей появления независимых кодовых символов.

Рассмотрим оптимальное эффективное кодирование сообщений источников без памяти и с памятью.

5.3.1. Кодирование сообщений источников без памяти. Символы сообщений источников без памяти являются независимыми. Покажем, чем определяется и как достигается минимальная длина кодовой комбинации при кодировании символов сообщений источников без памяти. Обеспечение минимальной средней длины кодовой комбинации и равновероятного появления в ней независимых кодовых символов равносильно увеличению средней скорости передачи информации до максимальной потому, что за одно и то же время коротких кодовых комбинаций можно передать больше, а это при прочих равных условиях соответствует передаче большего количества информации.

Минимальная средняя длина n_{0m} кодовой комбинации определяется совместным выполнением условий (5.29), (5.32):

$$n_{0m} \geqslant \frac{H_{\text{и}}(\text{B}_1)}{\log m_{\text{к}} + p_0 \log p_0 (m_{\text{к}} - 1)^{-1} + (1 - p_0) \log (1 - p_0)}. \qquad (5.42)$$

Если ошибки в канале отсутствуют, то минимальная средняя длина комбинации определяется только условием (5.29) при $H(\mathbf{B}_1|\mathbf{B}_2)=0$. Поэтому для идеальных каналов

$$n^*_{0m}=H_{\text{и}}(\mathbf{B}_1)/\log m_{\text{к}}. \qquad (5.43)$$

Например, при двоичном кодировании сообщений для идеального канала

$$n^*_{02}=H_{\text{и}}(\mathbf{B}_1). \qquad (5.44)$$

Следовательно, если энтропия двоичного кодера максимальна, среднее число кодовых символов в комбинациях минимально и равно энтропии источника. Оценка (5.44) средней длины кодовой комбинации является предельной; средняя длина кодовой комбинации не может быть меньше n^*_{02}. Эта оценка показывает, к чему необходимо стремиться при выборе способа эффективного кодирования. Теорема Шеннона для идеального канала утверждает, что такие способы кодирования существуют.

Если двоичный канал является реальным, то из-за ошибок растет средняя длина кодовых комбинаций и избыточность кода. Это вызвано необходимостью вводить в комбинации дополнительные кодовые символы, позволяющие обнаруживать и исправлять ошибки. В этом случае оценка средней минимальной длины кодовой комбинации подчиняется условию (5.40).

Для увеличения скорости передачи информации путем сокращения средней длины кодовых комбинаций символам сообщений, которые встречаются более часто, присваивают кодовые комбинации минимальной длины. Тогда короткие кодовые комбинации будут встречаться чаще и средняя длина кодовых комбинаций упадет. Код Морзе, например, построен по этому принципу, его недостаток в том, что приходится передавать разделительные символы — символы, обозначающие начало и конец кодовых комбинаций.

Этого недостатка лишен код Шеннона — Фано. Он построен по следующему алгоритму. Все символы алфавита сообщений записывают в порядке убывания вероятностей их появления. Полученную ранжированную (упорядоченную) последовательность символов разбивают на две группы так, чтобы суммы вероятностей групп были примерно одинаковыми. Для символов верхней группы в качестве первого символа кодовой комбинации присваивают кодовый символ 0, для символов нижней группы — кодовый символ 1. Полученные группы символов сообщений опять разбивают на две подгруппы по указанному принципу и опять кодируют. Такое разделение продолжают до тех пор, пока в последних подгруппах не останется по одному символу сообщений. Верхнему последнему символу в качестве последнего символа кодовой комбинации присваивается символ 0, нижнему — символ 1. При кодировании по этому алгоритму средняя длина кодовой комбинации близка к минимальной, энтропия кодера максимальна (вероятности появления кодовых символов 0 и 1 примерно одинаковы),

избыточность кода минимальна, скорость передачи информации близка к максимальной и приближается к пропускной способности канала.

Рассмотрим на конкретном примере особенности оптимального эффективного кодирования по алгоритму Шеннона-Фано. Предположим, что объем алфавита источника $m_и=5$, основание кода $m_к=2$, вероятности появления символов: $p_1=0{,}4$, $p_2=0{,}35$, $p_3=0{,}1$, $p_4=0{,}1$, $p_5=0{,}05$, канал идеальный. Оценим скорость передачи информации, пропускную способность и коэффициент использования идеального канала для равномерного кодирования и оптимального эффективного кодирования.

Вначале определим информационные характеристики источника сообщений. Энтропия источника

$$H_и(\mathbf{B}_1)=-\sum_{i=1}^{5} p_i \log p_i=1{,}94 \text{ бит/симв.},$$

максимальное значение энтропии $H_{и\,макс}=\log m_и=\log 5\simeq\simeq 2{,}32$ бит/симв., избыточность источника

$$r_и=1-H_и(\mathbf{B}_1)/\log m_и=1-1{,}94/2{,}32\simeq 0{,}164.$$

Найдем характеристики равномерного двоичного кода. Длину кодовых комбинаций определим из условия

$$m^n{}_к\geqslant m_и, \qquad (5.45)$$

согласно которому n необходимо выбирать так, чтобы общее число кодовых комбинаций было больше или равно числу символов алфавита сообщений. Использовав (5.45), получим, что ближайшим n, которое удовлетворяет условию (5.45), является $n_1=3$. Избыточность равномерного кода

$$r_{к1}=1-H_и(\mathbf{B}_1)/n_1\log m_к\simeq 1-1{,}94/3\simeq 0{,}353. \qquad (5.46)$$

Энтропия кодера

$$H_1(\mathbf{B}_1)=H_и(\mathbf{B}_1)/n_1=1{,}94/3\simeq 0{,}647 \text{ бит/сигн.}, \qquad (5.47)$$

скорость передачи информации $R_1=W_к H_1(\mathbf{B}_1)\simeq 0{,}647 W_к$ [бит/с], пропускная способность канала

$$C=W_к \log m_к=W_к. \qquad (5.48)$$

Коэффициент использования канала при равномерном кодировании

$$\eta_1=R_1/C=0{,}647 W_к/W_к=0{,}647. \qquad (5.49)$$

Найдем характеристики неравномерного двоичного кода Шеннона — Фано для этого случая. Особенности процедуры кодирования показаны в табл. 2 и с помощью графа кодирования на рис. 5.2. Граф кодирования показывает, как «расщепляется» ранжированная последовательность символов на группы и отдельные

символы и какие кодовые символы присваиваются группам и отдельным символам на каждом шаге разбиений.

Определим избыточность неравномерного кода

$$r_{к2}=1-H_и(\mathbf{B}_1)/n_{02}\log m_к=1-1,94/2\simeq 0,03.$$

Энтропия эффективного кодера

$$H_2(\mathbf{B}_1)=H_и(\mathbf{B}_1)/n_{02}=1,94/2=0,97, \qquad (5.50)$$

скорость передачи информации $R_2=W_кH_2(\mathbf{B}_1)=0,97W_к$, коэффициент использования канала $\eta_2=R_2/C=0,97W_к/W_к=0,97$.

Таблица 2

s_i	p_i	Порядковый номер разбиения				n_i	$n_i p_i$
		1	2	3	4		
s_1	0,4	0				1	0,4
s_2	0,35	1	0			2	0,7
s_3	0,1	1	1	0		3	0,3
s_4	0,1	1	1	1	0	4	0,4
s_5	0,05	1	1	1	1	4	0,2

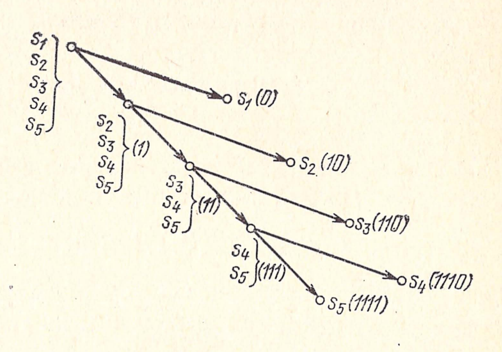


Рис. 5.2. Граф кодирования

Анализ и сравнение результатов равномерного и оптимального эффективного (неравномерного) кодирования для идеального канала в рассмотренном примере показывают следующее. Энтропия эффективного кодера и скорость передачи информации при неравномерном кодировании примерно на 33% выше, чем при равномерном, избыточность кода — на 33% ниже, коэффициент использования канала выше на 33% и близок к единице. Средняя длина n_{02} кодовых комбинаций (5.44) при неравномерном кодировании близка к минимальной: $n^*_{02}=1,94$. Так как максимальное количество информации, которое может переносить двоичный кодовый сигнал, равно 1 бит, то важно отметить, что энтропия эффективного кодера близка к максимальной — каждый кодовый сигнал несет 0,97 бит информации. Следовательно, в результате кодирования символы 0 и 1 появляются почти с одинаковой вероятностью. Таким образом, эффективный кодер преобразовал неравновероятные независимые символы источника сообщений в равновероятные независимые кодовые сигналы.

Важным свойством кода Шеннона — Фано является отсутствие характерных для других кодов трудностей в определении границ кодовых комбинаций. Коды, которые обладают такими свойствами, называют *неприводимыми*, они позволяют однозначно декодировать кодовые комбинации. Код Шеннона — Фано является неприводимым потому, что короткие комбинации никогда не явля-

ются началом более длинных кодовых комбинаций, благодаря этому не требуется разделительных знаков между кодовыми комбинациями.

Соотношение (5.48) может навести на мысль, что для двоичного канала пропускная способность равна скорости передачи сигналов во всех случаях. Это справедливо для тех случаев, когда все кодовые сигналы являются переносчиками информации. Когда есть сигналы, не несущие информации, скорость передачи сигналов и пропускная способность имеют различные значения. Например, стартстопный телеграфный аппарат передает один символ сообщения семью посылками: одной пусковой длительностью 20 мс, пятью кодовыми длительностью 20 мс каждая и одной стоповой длительностью 30 мс. Скорость передачи телеграфных сигналов

$$W_{\text{к}}=1/\tau=1/0{,}02=50 \text{ сигн./с}=50 \text{ Бод}, \qquad (5.51)$$

а пропускная способность определяется количеством только кодовых посылок в 1 с (пусковая и стоповая посылки, которые используют для синхронизации, информации не несут), поэтому

$$C=5 \text{ бит}/0{,}15 \text{ с}\simeq 33 \text{ бит/с}. \qquad (5.52)$$

Для сравнения минимальной длины кодовых комбинаций в идеальном и реальном каналах покажем, как изменится для рассмотренного примера значение n_{02}, если в канале есть ошибки и вероятность появления ошибки при передаче одного кодового сигнала $p_0=0{,}1$. В соответствии с (5.42) получим

$$n'_{02}>1{,}94/(1+0{,}1\log 0{,}1+0{,}9\log 0{,}9)=1{,}94/0{,}75=2{,}6 \text{ сигн./симв.}$$

Избыточность неравномерного кода для реального канала (см. (5.30), (5.41))

$$r_{\text{кз}}>1-H_{\text{и}}(\mathbf{B}_1)/n'_{02}\log m_{\text{к}}=1-1{,}94/2{,}6=0{,}25.$$

Энтропия кодера

$$H_{\text{з}}(\mathbf{B}_1)<H_{\text{и}}(\mathbf{B}_1)/n'_{02}=1{,}94/2{,}6=0{,}75 \text{ бит/сигн.}$$

Скорость передачи информации $R_{\text{з}}<W_{\text{к}}H_{\text{з}}(\mathbf{B}_1)=0{,}75W_{\text{к}}$ [бит/с], пропускная способность канала

$$C_{\text{з}}=W_{\text{к}}(1+0{,}1\log 0{,}1+0{,}9\log 0{,}9)=0{,}75W_{\text{к}} \text{ [бит/с]},$$

коэффициент использования канала $\eta_{\text{з}}<1$.

Следовательно, для обнаружения и исправления ошибок в канале и обеспечения сколь угодно малой вероятности ошибки избыточность кодирования должна быть повышена более чем на 25%. Энтропия кодера и скорость передачи информации при этом также уменьшается более чем на 25%. Интересно отметить, что избыточность, искусственно вводимая для коррекции ошибок в реальных каналах с высоким уровнем помех, может превысить ту естественную избыточность, которая устраняется при оптимальном эффективном кодировании.

5.3.2. Кодирование сообщений источников с памятью. Если источник имеет память на l_1 символов, то для устранения межсимвольной корреляции применяют укрупнение первичного алфавита. В роли «символов» вторичного алфавита выступают последовательности (блоки) из символов первичного алфавита. В результате укрупнения обеспечивается переход от коррелированных символов первичного алфавита к независимым блокам вторичного алфавита. После такой перекодировки независимые неравновероятные блоки кодируют способами, применяемыми для источников без памяти.

Следовательно, кодирование сообщений источников с памятью выполняют в два этапа: на первом этапе осуществляют разбивку сообщений на блоки длиной $l>l_1$, в результате чего блоки нового алфавита становятся независимыми, а на втором этапе используют оптимальные коды, подобные коду Шеннона — Фано.

Рассмотрим особенности перекодировки зависимых символов первичного алфавита в независимые блоки вторичного. Кодирование блока длиной l символов можно начать лишь тогда, когда он полностью поступил на декодер. Декодирование также может начаться только после того, как принят весь блок. Поэтому время передачи одного блока

$$\tau=2l\tau_1+\tau_0, \tag{5.53}$$

где τ_1 — время передачи одного символа первичного алфавита; τ_0 — задержка блоков в канале. Производительность источника блоков

$$H'_\text{и}(l)=H_\text{и}(l)\,/\,(2l\tau_1+\tau_0),$$

где $H_\text{и}(l)$ — среднее количество информации в одном блоке.

Укрупнение алфавита не изменяет избыточности сообщений. Однако избыточность, обусловленная корреляционными связями символов первичного алфавита, преобразуется в избыточность источника блоков, обусловленную неравновероятным появлением независимых блоков. Это объясняется тем, что неравномерность распределения вероятностей появления блоков больше, чем неравномерность распределения вероятностей появления символов первичного алфавита. Избыточность сообщений, составленных из блоков,

$$r_l=1-\frac{H_\text{и}(l)}{\log m_{\text{и}2}}=1-\frac{lH_\text{и}(\text{B}_1)}{\log m^l_\text{и}}=1-\frac{H_\text{и}(\text{B}_1)}{\log m_\text{и}}, \tag{5.54}$$

где $m_{\text{и}2}=m^l_\text{и}$ — объем укрупненного алфавита; $m_\text{и}$ — объем первичного алфавита. Из формулы (5.54) следует, что избыточность сообщений, составленных из символов первичного и вторичного алфавита, одна и та же.

Оптимальное эффективное кодирование сообщений почти полностью устраняет их избыточность. Из-за этого процесс передачи информации становится чувствительным к воздействию помех. Особенно сильно проявляется эта особенность при кодировании

сообщений источников с памятью. Ошибки в каналах могут привести к неправильному декодированию многих блоков, и следовательно, увеличение скорости передачи информации достигается за счет снижения верности. Поэтому оптимальное эффективное кодирование может быть использовано только для тех каналов, которые близки к идеальным.

Контрольные вопросы

1. В чем сущность оптимального эффективного кодирования?
2. Чем определяется минимальная средняя длина кодовых комбинаций при эффективном кодировании?
3. Какая идея лежит в основе эффективного кодирования?
4. Поясните алгоритм кодирования Шеннона — Фано.
5. Сравните информационные характеристики равномерного и неравномерного кода.
6. Как влияют ошибки в канале на информационные характеристики эффективного кода?
7. Укажите особенности эффективного кодирования сообщений источников с памятью.

5.4. ИНФОРМАЦИОННЫЕ ХАРАКТЕРИСТИКИ ИСТОЧНИКОВ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

Основные информационные характеристики источников непрерывных сообщений следующие: энтропия, условная энтропия, эпсилон-энтропия, эпсилон-производительность, избыточность, объем информации.

Формулу для энтропии источника непрерывных сообщений получают путем предельного перехода из формулы (1.8) для энтропии дискретного источника. Если одномерная плотность распределения случайного стационарного процесса $X(t)$ равна $f(x)$, то вероятность того, что $X(t)$ будет находиться в интервале $[x_i, x_i+\Delta x]$, равна $f(x_i)\Delta x$. Если всего интервалов квантования m, то энтропия источника непрерывных сообщений

$$H_{\Delta x} = -\sum_{i=1}^{m} f(x_i)\,\Delta x \log [f(x_i)\,\Delta x].$$

В результате предельного перехода при $\Delta x \to 0$ получим

$$\lim_{\Delta x \to 0} H_{\Delta x} = -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx - \lim_{\Delta x \to 0} \log \Delta x.$$

Первую составляющую

$$H_x = -\int_{-\infty}^{\infty} f(x) \log f(x)\,dx \qquad (5.55)$$

называют *дифференциальной энтропией источника* непрерывных сообщений. Вторая составляющая $\lim_{\Delta x \to 0} \log \Delta x \to \infty$ и показывает

что энтропия источника непрерывных сообщений стремится к бесконечности.

Скорость передачи информации, пропускную способность и другие основные информационные характеристики источников определяют через разность энтропий. Поэтому вторая составляющая в этих операциях сокращается и величиной H_x можно характеризовать информационные свойства источников непрерывных сообщений. В отличие от энтропии источников дискретных сообщений H_x может принимать положительные, отрицательные и нулевые значения. Величина H_x изменяется при изменении масштаба измерения $X(t)$. Определим свойства дифференциальной энтропии для конкретных распределений.

5.4.1. Энтропия равномерного распределения. Для равномерного распределения $f(x)=(b-a)^{-1}$ по формуле (5.55) получим

$$H_x=\log(b-a).$$

С увеличением интервала $b-a$ значение H_x также увеличивается. При отсутствии ограничений на значение дисперсии равномерное распределение обладает максимальной энтропией среди всех других распределений.

5.4.2. Энтропия нормального распределения. Определим плотность распределения центрированной случайной величины, обеспечивающую максимум H_x при условии выполнения нормировки и ограниченности дисперсии процесса.

Необходимо решить вариационную задачу в следующей постановке: заданы целевой функционал $H_x=-\int\limits_{-\infty}^{\infty} f(x)\log f(x)\,dx$ и ограничения $\int\limits_{-\infty}^{\infty} f(x)\,dx=1$, $\int\limits_{-\infty}^{\infty} x^2 f(x)\,dx=\sigma^2$. Необходимо найти такую $f(x)$, которая доставляет максимум H_x.

Решение этой задачи известно [1—3], искомая плотность распределения является гауссовской

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-x^2/2\sigma^2},$$

а максимальная энтропия

$$H_{x\,\text{макс}}=\log\sqrt{2\pi e\sigma^2}. \tag{5.56}$$

Следовательно, среди всех источников с ограниченной и одинаковой мощностью непрерывных сигналов наибольшей энтропией обладает источник с гауссовскими сигналами; максимальную энтропию полностью определяет мощность σ^2 сигнала; от среднего значения сигнала энтропия не зависит.

5.4.3. Энтропию гауссовского белого шума определим как энтропию n-мерного гауссовского распределения с независимыми отсчетами, тогда

$$H_n=\log(\sqrt{2\pi e\sigma^2})^n=nH_{x\,\text{макс}}. \tag{5.57}$$

Следовательно, энтропия гауссовского белого шума в n раз больше энтропии одного отсчета и является максимальной для непрерывных сигналов ограниченной мощности.

Из анализа (5.57) можно сделать следующие выводы: гауссовский белый шум, обладая наибольшей энтропией, является наиболее вредной помехой, так как переносит максимально возможное количество вредной информации при заданной средней мощности; наихудшие («пессимистические») оценки качества связи получаются, если в роли помех использовать белый шум; если в роли сигналов-переносчиков использовать гауссовские сигналы, по свойствам близкие к белому шуму, можно при одной и той же средней мощности передать максимальное количество информации. Иначе говоря, применение шумоподобных сигналов-переносчиков дает возможность передать максимум информации при заданной мощности или уменьшить мощность сигналов при передаче заданного количества информации. Из этих выводов понятно, почему белый шум широко используют как модель сигналов-переносчиков и помех.

5.4.4. Эпсилон-энтропия источников. Реальная чувствительность приемных устройств, органов чувств человека и разрешающая способность различных информационно-измерительных систем ограничены. Поэтому воспроизводить непрерывные сообщения абсолютно точно не требуется. Наличие помех и искажений сигналов в реальных каналах делает точное воспроизведение сообщений невозможным. Поэтому вводят понятие эпсилон-энтропии. Эпсилон-энтропия — это то среднее количество информации в одном независимом отсчете непрерывного случайного процесса $X(t)$, которое необходимо для воспроизведения этого сигнала с заданной среднеквадратической погрешностью ε_0.

Рассмотрим подробнее сущность этого понятия. Предположим, что передавался сигнал $X(t)$, а был принят сигнал $Y(t)$. Пусть в канале действует аддитивная помеха $\xi(t)$, тогда $Y(t)=X(t)+\xi(t)$. Условно примем, что $M[\xi(t)]=0$, $M[\xi^2(t)]=\varepsilon^2$. Используем метрику гильбертова пространства (см. § 2.10), тогда расстояние между сигналами $X(t)$ и $Y(t)$ определяется величиной

$$\varepsilon^2=\frac{1}{T}\int_0^T [Y(t)-X(t)]^2\,dt=\frac{1}{T}\int_0^T \xi^2(t)\,dt, \qquad (5.58)$$

где T — длительность сигналов. Если $\varepsilon^2 \leqslant \varepsilon^2_0$, то сигналы называют ε_0-близкими.

Используем ортогональное разложение Котельникова (см. § 2.2) и выполним дискретизацию сигналов. Тогда вместо непрерывных сигналов $X(t)$ и $Y(t)$ будем рассматривать отсчеты $\overline{X_1, X_n}$ и $\overline{Y_1, Y_n}$. Введем следующие обозначения: $f_1(\overline{x_1, x_n})=f_1(\mathbf{X})$ — n-мерная плотность распределения $X(t)$, $f_2(\mathbf{Y})$ — n-мерная плотность распределения $Y(t)$, $f_1(\mathbf{X}|\mathbf{Y})$ — условная плотность распределения $\mathbf{X}$ при условии, что был принят $\mathbf{Y}$, $f_2(\mathbf{Y}|\mathbf{X})$ — условная плотность распределения $\mathbf{Y}$ при условии, что был передан $\mathbf{X}$.

Количество информации, которое содержит принятый сигнал $\mathbf{Y}$ относительно переданного сигнала $\mathbf{X}$, определяется, как и ранее (см. п. 5.2.2):

$$I_1(\mathbf{Y},\ \mathbf{X}) = \log \frac{f_1(\mathbf{X}\,|\,\mathbf{Y})\, f_2(\mathbf{Y})}{f_1(\mathbf{X})\, f_2(\mathbf{Y})} = \log \frac{f_1(\mathbf{X}\,|\,\mathbf{Y})}{f_1(\mathbf{X})} = \log \frac{f_2(\mathbf{Y}\,|\,\mathbf{X})}{f_2(\mathbf{Y})}. \qquad (5.59)$$

Величина (5.59) является аналогом величины $I(b_{\text{к}2},\ b_{i1})$ (5.20) при передаче дискретных сигналов.

Среднее количество принятой информации

$$I(\mathbf{Y},\ \mathbf{X}) = \int\limits_{\mathbf{X}} \int\limits_{\mathbf{Y}} I_1(\mathbf{Y},\ \mathbf{X})\, f(\mathbf{Y},\ \mathbf{X})\, d\mathbf{X} d\mathbf{Y}, \qquad (5.60)$$

где интегралы являются n-мерными; совместная плотность распределения $\mathbf{Y}$ и $\mathbf{X}$

$$f(\mathbf{Y},\ \mathbf{X}) = f_1(\mathbf{X}) f_2(\mathbf{Y}\,|\,\mathbf{X}) = f_2(\mathbf{Y}) f_1(\mathbf{X}\,|\,\mathbf{Y}). \qquad (5.61)$$

Используя (5.61), получаем

$$I(\mathbf{Y},\ \mathbf{X}) = H_{\mathbf{X}} - H_{\mathbf{X}|\mathbf{Y}}, \qquad (5.62)$$

$$I(\mathbf{Y},\ \mathbf{X}) = H_{\mathbf{Y}} - H_{\mathbf{Y}|\mathbf{X}}. \qquad (5.63)$$

Соотношения (5.62), (5.63) аналогичны соотношениям (5.24), (5.25) для случая передачи дискретных сигналов. Энтропия $\mathbf{X}$ (5.55)

$$H_{\mathbf{X}} = \int\limits_{\mathbf{X}} f_1(\mathbf{X}) \log f_1(\mathbf{X})\, d\mathbf{X}, \qquad (5.64)$$

условная энтропия

$$H_{\mathbf{X}|\mathbf{Y}} = -\int\limits_{\mathbf{X}} \int\limits_{\mathbf{Y}} f_1(\mathbf{X}\,|\,\mathbf{Y}) \log f_1(\mathbf{X}\,|\,\mathbf{Y})\, d\mathbf{X} d\mathbf{Y}. \qquad (5.65)$$

Условная энтропия $H_{\mathbf{X}|\mathbf{Y}}$ источников непрерывных сигналов обладает теми же свойствами, что и условная энтропия источника дискретных сигналов:

$$H_{\mathbf{X}|\mathbf{Y}} < H_{\mathbf{X}}, \quad H_{\mathbf{XY}} = H_{\mathbf{X}} + H_{\mathbf{Y}|\mathbf{X}}, \quad H_{\mathbf{XY}} \leqslant H_{\mathbf{X}} + H_{\mathbf{Y}}, \qquad (5.66)$$

где $H_{\mathbf{XY}}$ — совместная энтропия $\mathbf{X}$ и $\mathbf{Y}$. Только в том случае, когда сигналы независимы, условная энтропия равна безусловной, а совместная энтропия сигналов равна сумме энтропий.

Как и для дискретных источников, среднее количество принятой информации можно определить двумя способами: как разность среднего количества переданной информации $H_{\mathbf{X}}$ и информации $H_{\mathbf{X}|\mathbf{Y}}$, потерянной в непрерывном канале из-за помех (5.62), и как разность между средним количеством принятой информации $H_{\mathbf{Y}}$ и энтропией помехи в канале (5.63). Заметим, что слагаемые вида

$\lim\limits_{\Delta X \to 0} \log \Delta X$ при определении $I_1(\mathbf{X}, \mathbf{Y})$, $I(\mathbf{Y}, \mathbf{X})$ сокращаются и на результат влияния не оказывают.

Эпсилон-энтропия $H_{\mathbf{X}\varepsilon}$ — это то минимальное среднее количество информации $\min I(\mathbf{Y}, \mathbf{X})$, которое позволяет считать $Y(t)$ и $X(t)$ «похожими» в среднеквадратическом смысле:

$$H_{\mathbf{X}\varepsilon} = \min I(\mathbf{Y}, \mathbf{X}) = H_{\mathbf{X}} - \max_{f_1(\mathbf{X}|\mathbf{Y})} H_{\mathbf{X}|\mathbf{Y}}. \quad (5.67)$$

Так как $X(t) = Y(t) - \xi(t)$, то условная энтропия $H_{\mathbf{X}|\mathbf{Y}}$ при принятом $Y(t)$ полностью определяется «шумом» воспроизведения $\xi(t)$. Поэтому

$$\max_{f_1(\mathbf{X}|\mathbf{Y})} H_{\mathbf{X}|\mathbf{Y}} = \max_{f_3(\xi)} H_\xi, \quad (5.68)$$

где $f_3(\xi)$ — закон распределения помехи $\xi(t)$.

Учтем, что мощность помехи ограничена величиной ε^2_0, тогда максимальная энтропия помехи, отнесенная к одному отсчету, определяется по формуле (5.56)

$$\max_{f_3(\xi)} H_\xi = \log \sqrt{2\pi e}\, \varepsilon_0 = \sigma_\xi \log \sqrt{2\pi e}, \quad (5.69)$$

где через σ_ξ обозначено среднеквадратическое значение помехи. С учетом (5.69)

$$H_{X\varepsilon_0} = H_X - \log \sqrt{2\pi e}\, \varepsilon_0. \quad (5.70)$$

Эпсилон-энтропия имеет максимальное значение, когда процесс $X(t)$ также является гауссовским:

$$\max_{f_1(X)} H_{X\varepsilon_0} = \log \sqrt{2\pi e}\, \sigma_x - \log \sqrt{2\pi e}\, \sigma_\xi = 0{,}5 \log (\sigma^2_X / \sigma^2_\xi). \quad (5.71)$$

Отношение сигнал/шум $\sigma^2_X / \sigma^2_\xi$ характеризует то количество полученной информации, при котором принятый сигнал $Y(t)$ и переданный сигнал $X(t)$ «похожи» в среднеквадратическом смысле с точностью до $\varepsilon^2_0 = \sigma^2_\xi$. В формуле (5.71) значение эпсилон-энтропии определено для одного независимого отсчета.

5.4.5. Эпсилон-производительность источника. Если источник выдает независимые отсчеты сигнала $X(t)$ в дискретные моменты времени со скоростью $W_и = 1/\Delta t$, где интервал дискретизации $\Delta t = 1/2\Delta F$ (ΔF — полоса частот сигнала $X(t)$), то эпсилон-производительность источника

$$H'_{X\varepsilon_0} = W_и H_{X\varepsilon_0} = W_и (H_X - \log \sqrt{2\pi e\, \varepsilon^2_0}) \text{ [бит/с]}. \quad (5.72)$$

Если время непрерывное, то

$$H'_{X\varepsilon_0} = 2\Delta F [H_X - \log \sqrt{2\pi e \varepsilon^2_0}] \text{ [бит/с]}. \quad (5.73)$$

Максимальное значение эпсилон-производительность источника имеет, когда сигнал $X(t)$ является гауссовским (см. (5.71)),

$$\max_{f_1(X)} H'_{X\varepsilon_0} = \frac{W_{\text{и}}}{2} \log(\sigma^2_X/\varepsilon^2_0) \text{ [бит/с]},$$

$$\max_{f_1(X)} H'_{X\varepsilon_0} = F \log(\sigma^2_X/\varepsilon^2_0) \text{ [бит/с]}.$$

За время T существования сигнала максимальный объем V информации, выданной источником,

$$\max_{f_1(X)} V = \max_{f_1(X)} H'_{X\varepsilon_0} T = FT \log(\sigma^2_X/\varepsilon^2_0) \text{ [бит]}. \quad (5.74)$$

Поэтому объем сигнала является одновременно экстремальной информационной характеристикой сигнала. *Объем сигнала* — это максимальное количество информации, которое сигнал может переносить.

5.4.6. Избыточность источника непрерывных сигналов определяют так же, как и для источника дискретных сигналов:

$$r_X = 1 - \frac{H_{X\varepsilon_0}}{\max H_{X\varepsilon_0}} = 1 - \frac{H_X - \log\sqrt{2\pi e\varepsilon^2_0}}{0{,}5\log(\sigma^2_X/\varepsilon^2_0)}. \quad (5.75)$$

Избыточность источника равна нулю только в случае, когда распределение сигнала является гауссовским.

При определении эпсилон-характеристик источников непрерывных сигналов в качестве меры близости двух сигналов было выбрано расстояние между ними в гильбертовом пространстве. Критерием близости служило среднеквадратическое отклонение одного сигнала от другого. Если выбрать другую меру близости сигналов — другую метрику пространства сигналов, можно получить другие эпсилон-характеристики источников. Наибольшее распространение получил среднеквадратический критерий близости сигналов.

Контрольные вопросы

1. Укажите основные информационные характеристики источников непрерывных сообщений.
2. Как определяют энтропию непрерывного источника?
3. Какие распределения сигналов имеют максимальную энтропию?
4. Чему равна энтропия гауссовского сигнала, гауссовского белого шума?
5. Что такое эпсилон-энтропия?
6. Как определяют среднее количество принятой информации для непрерывных источников?
7. Как определяют эпсилон-энтропию, эпсилон-производительность и избыточность источника непрерывных сообщений?
8. Как связаны объем сигнала и количество переносимой им информации?

Основными информационными характеристиками непрерывного канала, так же как и дискретного, являются скорость передачи информации, пропускная способность и коэффициент использования.

5.5.1. Скорость передачи информации и пропускная способность. Если $X(t)$ рассматривать как переданный сигнал, $Y(t)$ — как принятый, а $\xi(t)$ — как аддитивную помеху в непрерывном канале (см. § 4.1), то *скорость передачи информации по непрерывному каналу*

$$R = W_{и} [H_X - H_{X|Y}] = W_{и} [H_Y - H_{Y|X}], \tag{5.76}$$

где H_X определяется соотношением (5.64), а $H_{X|Y}$ и $H_{Y|X}$ — соотношениями типа (5.65).

Пропускная способность непрерывного канала без памяти

$$C = \max R = W_{и} \max [H_X - H_{X|Y}] = W_{и} \max [H_Y - H_{Y|X}]. \tag{5.77}$$

Для гауссова непрерывного канала с дискретным временем $C = W_{и} \max [H_Y - H_\xi]$. Учтем, что $\max\limits_{f_2(Y)} H_Y = \log \sqrt{2\pi e (\sigma^2_X + \sigma^2_\xi)}$,

$$\max_{f_3(\xi)} H_\xi = \log \sqrt{2\pi e \sigma^2_\xi},$$

тогда

$$C = 0{,}5 W_{и} \log (1 + \sigma^2_X / \sigma^2_\xi). \tag{5.78}$$

Если в канале нет искажений и помех, то σ^2_ξ можно рассматривать как мощность шумов квантования при дискретной передаче непрерывных сигналов. В канале с помехами мощность шумов квантования складывается с мощностью помех, следовательно, в этом случае σ^2_ξ необходимо рассматривать как суммарную мощность помехи и шума квантования.

Для непрерывного канала с непрерывным временем $W_{и} = 2\Delta F$ и формула (5.78) переходит в известную формулу Шеннона для пропускной способности гауссова непрерывного канала с флуктуационной помехой:

$$C = \Delta F \log (1 + \sigma^2_X / \sigma^2_\xi). \tag{5.79}$$

Эта формула связывает полосу сигнала, отношение сигнал/шум и пропускную способность непрерывного канала. Из нее следует, что одну и ту же пропускную способность можно получить при различных соотношениях ΔF и $\sigma^2_X / \sigma^2_\xi$. Пропускная способность линейно зависит от полосы сигнала, поэтому более целесообразно обменивать мощность сигнала на полосу и заданную пропускную способность обеспечивать за счет применения широкополосных (шумоподобных) сигналов.

Рассмотрим предельные характеристики пропускной способности при изменении ΔF. Так как спектральная плотность белого шума $\Omega_0 = \sigma^2_\xi / \Delta F$, то удобно ввести эквивалентную полосу шума $\Delta F_0 = \sigma^2_X / \Omega_0$ (ΔF_0 — полоса спектра такого шума, который при спектральной плотности Ω_0 имеет мощность, равную мощности сигнала). Подставив ΔF_0 в (5.79), получим $C = \Delta F \log(1 + \Delta F_0 / \Delta F)$. Пронормируем пропускную способность по эквивалентной полосе шума:

$$C / \Delta F_0 = (\Delta F / \Delta F_0) \log(1 + \Delta F_0 / \Delta F).$$

Определим

$$\lim_{\Delta F \to \infty} \frac{C}{\Delta F_0} = \lim_{F \to \infty} \frac{\Delta F}{\Delta F_0} \log\left(1 + \frac{\Delta F_0}{\Delta F}\right) = \log e = C_\infty = 1{,}443 \text{ бит}.$$

На рис. 5.3 показан график зависимости нормированной безразмерной пропускной способности $C_0 = C / \Delta F_0 C_\infty$ от отношения $\Delta F / \Delta F_0$. Из анализа графика следует, что чем шире спектр сигнала (чем больше сигнал приближается к белому шуму), тем больше пропускная способность приближается к предельной. В этом одно из основных преимуществ применения шумоподобных (широкополосных) сигналов.

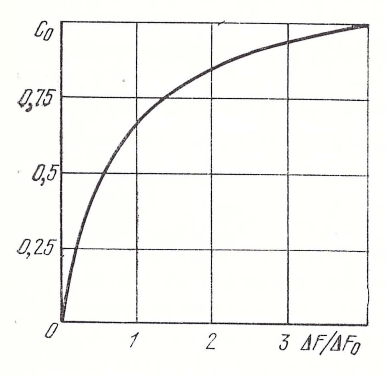


Рис. 5.3. График $C_0(\Delta F/\Delta F_0)$

5.5.2. Пропускная способность каналов с замираниями. В каналах с замираниями модуль коэффициента передачи $\mathscr{K}$ и фаза φ сигнала являются случайными величинами (см. п. 4.1.3, (4.10), (4.11)). Мощность выходного сигнала равна $\mathscr{K}^2 \sigma^2_x$, а средняя мощность — $\mathscr{K}^2_0 \sigma^2_x$, где $\mathscr{K}_0$ — среднее значение модуля коэффициента передачи.

Пропускную способность канала с замираниями определяют как функцию случайного аргумента $\mathscr{K}$ по известному распределению $f(\mathscr{K})$. Математическое ожидание пропускной способности

$$M[C] = \int_0^\infty \Delta F \log\left(1 + \frac{\mathscr{K}^2 \sigma^2_x}{\mathscr{K}^2_0 \sigma^2_\xi}\right) f(\mathscr{K})\, d\mathscr{K}. \tag{5.80}$$

Например, если замирания релеевские (см. п. 4.1.3) и пропускная способность исчисляется в натуральных единицах в секунду [9], то

$$M[C_1] = \int_0^\infty \Delta F \ln\left(1 + \frac{\mathscr{K}^2 \sigma^2_x}{\mathscr{K}^2_0 \sigma^2_\xi}\right) f(\mathscr{K}) d\mathscr{K} =$$

$$= \int_0^\infty \Delta F \ln\left(1 + \frac{\mathscr{K}^2 \sigma^2_x}{\mathscr{K}^2_0 \sigma^2_\xi}\right) \frac{2\mathscr{K}}{\sigma^2_k} e^{-\mathscr{K}^2/\sigma^2_k}\, d\mathscr{K} \approx -\Delta F \exp\frac{\sigma^2_\xi}{\sigma^2_x} \operatorname{Ei}\left(\frac{\sigma^2_\xi}{\sigma^2_x}\right),$$

где $\mathrm{Ei}(x) = \int\limits_{-\infty}^{x} \frac{e^t}{t} dt$ — табулированная интегральная показательная функция [9].

Если $\sigma^2{}_x/\sigma^2{}_\xi \gg 1$, то

$$M[C_2] \cong \Delta F \left(1 + \frac{\sigma^2{}_\xi}{\sigma^2{}_x}\right)\left(\ln \frac{\sigma^2{}_x}{\sigma^2{}_\xi} - 0,5772\right).$$

Если $\sigma^2{}_x/\sigma^2{}_\xi \ll 1$, то [9]

$$M[C_3] \cong \Delta F \sigma^2{}_x/\sigma^2{}_\xi \cong \sigma^2{}_x/\Omega_0.$$

Релеевские замирания снижают пропускную способность канала не более чем на 17%. Пропускную способность каналов с другими видами замираний определяют аналогично.

5.6. СОГЛАСОВАНИЕ ИСТОЧНИКОВ С КАНАЛАМИ

Предельные возможности согласования источника дискретных сообщений с непрерывным каналом определяются *теоремой кодирования Шеннона*: если производительность $H'_\text{и}(\mathbf{B}_1)$ источника дискретных сообщений меньше пропускной способности непрерывного канала, то существует способ оптимального кодирования и декодирования, при котором вероятность ошибки сколь угодно мала. При $H'_\text{и}(\mathbf{B}_1) > C$ такого способа нет. Под оптимальным кодированием сообщений для передачи по непрерывному каналу понимают преобразование длинных последовательностей символов источника в длинные непрерывные сигналы. Этот способ отличается от посимвольного кодирования, так как здесь кодируют последовательности символов большой длины.

Как следует из теоремы Шеннона, для гауссова непрерывного канала скорость кодирования

$$W_\text{к} < \frac{\Delta F_\text{к} \log(1 + \sigma^2{}_x/\sigma^2{}_\xi)}{H(\mathbf{B}_1)}, \tag{5.81}$$

где $\Delta F_\text{к}$ — полоса канала. Если кодер выдает равновероятные и независимые двоичные сигналы, то $H(\mathbf{B}_1) = H_\text{макс} = 1$ и

$$W_\text{к} < \Delta F_\text{к} \log(1 + \sigma^2{}_x/\sigma^2{}_\xi). \tag{5.82}$$

Следовательно, при больших отношениях сигнал/шум скорость кодирования для непрерывного канала может превышать предел Найквиста $2\Delta F_\text{к}$.

Для того чтобы оценить, как используется пропускная способность непрерывных каналов при передаче дискретных сообщений, вводят *коэффициент эффективности передачи дискретных сообщений* η_0, который определяют как отношение пропускной способ-

ности дискретного канала к пропускной способности непрерывного канала

$$\eta_0 = \frac{W_К [\log m_К + p_0 \log p_0 (m_К - 1)^{-1} + (1 - p_0) \log (1 - p_0)]}{\Delta F_К \log(1 + \sigma^2_x / \sigma^2_\xi)}. \quad (5.83)$$

При оптимальном многопозиционном кодировании в принципе можно обеспечить близкие к единице значения η_0 (см. § 5.6), однако техническая реализация такого кодирования сложна.

Предельные возможности согласования источника непрерывных сообщений с непрерывным каналом определяются следующей *теоремой кодирования Шеннона*: если эпсилон-производительность $H'_{X\varepsilon_0}$ источника непрерывных сообщений меньше пропускной способности канала, то существует способ оптимального кодирования и декодирования, при котором с вероятностью, сколь угодно близкой к единице, переданное и принятое сообщения не будут отличаться в среднеквадратическом смысле более чем на ε^2_0.

При $H'_{X\varepsilon_0} > C$ такого способа нет. Под оптимальным кодированием непрерывных сообщений в непрерывные сигналы понимают преобразование без предварительной дискретизации по времени и квантования по уровню. Речь идет о выборе способа аналоговой модуляции, оптимальное кодирование соответствует идеальной модуляции.

Для гауссова канала условие существования оптимального кодирования принимает вид

$$\Delta F \log (\sigma^2_c / \varepsilon^2_0) < \Delta F_К \log (1 + \sigma^2_c / \sigma^2_\xi), \quad (5.84)$$

где отношение $\sigma^2_c / \varepsilon^2_0$ рассматривается на выходе канала (для копии сообщения), а отношение $\sigma^2_c / \sigma^2_\xi$ — в канале. Например, величину $\sigma^2_c / \varepsilon^2_0$ можно рассматривать как отношение сигнал/шум на выходе детектора, а величину $\sigma^2_c / \sigma^2_\xi$ — как отношение сигнал/шум на входе приемника.

Как следует из (5.84), необходимое значение $\sigma^2_c / \varepsilon^2_0$ можно обеспечить в узкополосном канале:

$$\Delta F_К / \Delta F \approx 1, \quad (5.85)$$

и в широкополосном, для которого $\Delta F_К / \Delta F \gg 1$. Чтобы это показать, представим (5.84) в следующем виде:

$$\sigma^2_c / \varepsilon^2_0 < (1 + \sigma^2_c / \sigma^2_\xi)^{\Delta F_К / \Delta F}. \quad (5.86)$$

Если канал узкополосный, необходимо обеспечивать очень большие отношения $\sigma^2_c / \sigma^2_\xi$.

Коэффициент η_1 *эффективности передачи непрерывных сигналов* по непрерывному каналу определяют как η_0 (5.83). Для гауссова непрерывного источника и канала без памяти

$$\eta_1 = \frac{\Delta F}{\Delta F_К} \frac{\log [1 + (\mathcal{P}_c / \mathcal{P}_ш)_2]}{\log [1 + (\mathcal{P}_c / \mathcal{P}_ш)_1]}, \quad (5.87)$$

где индексом 2 обозначено отношение сигнал/шум на выходе канала (на выходе приемника), а индексом 1 — в канале (на входе в приемник). Для сокращения записи введены мощности сигнала $\mathscr{P}_c=\sigma^2_c$ и помехи (шума) $\mathscr{P}_ш=\sigma^2_\xi$. По существу η_1 характеризует эффективность способа модуляции. Для идеальной модуляции $\eta_1=1$ и

$$(\mathscr{P}_c/\mathscr{P}_ш)_2=[1+(\mathscr{P}_c/\mathscr{P}_ш)_1]^{\Delta F_к/\Delta F}-1. \qquad (5.88)$$

Для реальных непрерывных систем часто используют *коэффициент эффективности модуляции*

$$\mathfrak{p}=\frac{(\mathscr{P}_c/\mathscr{P}_ш)_2}{(\mathscr{P}_c/\mathscr{P}_ш)_1},$$

называемый *выигрышем модуляции*. Для *идеальной системы*

$$\mathfrak{p}_0=\frac{[1+(\mathscr{P}_c/\mathscr{P}_ш)_1]^{\Delta F_к/\Delta F}-1}{(\mathscr{P}_c/\mathscr{P}_ш)_1}. \qquad (5.89)$$

Если $\Delta F_к/\Delta F=1$, что имеет место при однополосной модуляции, то при любых отношениях $(\mathscr{P}_c/\mathscr{P}_ш)_1 \mathfrak{p}_0=1$. Если $\Delta F_к/\Delta F>1$, то $\mathfrak{p}_0>1$ и растет с ростом отношения $(\mathscr{P}_c/\mathscr{P}_ш)_1$.

Контрольные вопросы

1. Назовите основные информационные характеристики непрерывного канала.
2. Как определяют скорость передачи информации и пропускную способность непрерывного канала?
3. Как зависит пропускная способность непрерывного канала от ширины спектра сигнала?
4. Сформулируйте две основные теоремы Шеннона об оптимальном кодировании для непрерывного канала.
5. Чему равны коэффициенты эффективности передачи дискретных и непрерывных сигналов?
6. Как определяют эффективность модуляции? Чему равна эффективность модуляции для идеальной системы?

5.7. СРАВНЕНИЕ ПРОПУСКНЫХ СПОСОБНОСТЕЙ ДИСКРЕТНОГО И НЕПРЕРЫВНОГО КАНАЛОВ

Сравнение нормированных по полосе частот канала пропускных способностей каналов выполним для гауссова непрерывного канала и построенного на его основе двоичного симметричного дискретного канала без памяти. Нормированная пропускная способность гауссова непрерывного канала определяется формулой (5.79) при условии, что $\Delta F=\Delta F_к$

$$\frac{C_1}{\Delta F_к}=\log\left(1+\frac{\mathscr{P}_c}{\mathscr{P}_ш}\right) \text{ [бит]}. \qquad (5.90)$$

Нормированная пропускная способность двоичного симметричного канала, построенного на этом непрерывном канале, определяется формулой (5.38)

$$C_2/\Delta F_к=2[1+p_0\log p_0+(1-p_0)\log(1-p_0)]. \qquad (5.91)$$

На рис. 5.4 показаны графики зависимостей (5.90) и (5.91) от отношения сигнал/шум. Можно заметить, что предельное значение нормированной пропускной способности двоичного дискретного канала, равное двум битам, достигается при $\mathscr{P}_с/\mathscr{P}_ш=4-5$, когда $p_0 \to 0$, а выражение в квадратных скобках формулы (5.91) стремится к единице. Нормированная пропускная способность непрерывного канала является монотонно возрастающей неограниченной функцией отношения сигнал/шум. Оценки нормированных пропускных способностей двоичного и непрерывного каналов близки при малых значениях отношения сигнал/шум ($\mathscr{P}_с/\mathscr{P}_ш=$ $=0{,}2-2$).

Для многопозиционных дискретных каналов характер изменения нормированной пропускной способности такой же, изменяется только предельное значение нормированной пропускной способности, равное

$$\lim_{h\to\infty}(C_m/\Delta F_к)=2\log m_к,$$

где $h=\mathscr{P}_с/\mathscr{P}_ш$. Следовательно, пропускную способность дискретного канала можно увеличить путем увеличения числа позиций сигналов. При $m_к \to \infty$ нормированная пропускная способность дискретного канала стремится к пропускной способности непрерывного канала. Это объясняется тем, что при $m_к \to \infty$ сигналы на выходе дискретного модулятора приближаются по своим свойствам к гауссовскому белому шуму.

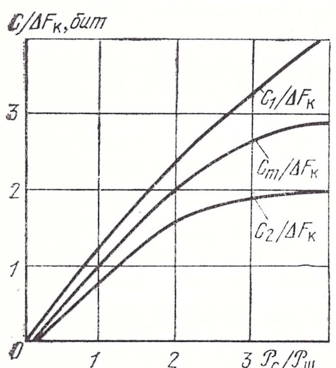


Рис. 5.4. Сравнение пропускных способностей дискретных и непрерывного каналов

Следует отметить влияние сосредоточенных помех на пропускную способность. Так как гауссовский белый шум является наиболее тяжелым видом помех, то при всех других видах помех пропускная способность будет большей. Если одновременно действуют сосредоточенные помехи и флуктуационная помеха, то пропускная способность, естественно, будет меньше. Если непрерывный канал имеет переменные параметры (см. п. 5.5.2) или сигналы подвергаются линейным и нелинейным искажениям, то средняя мощность сигналов на выходе канала, отношение сигнал/шум и, следовательно, пропускная способность канала меньше. Рассмотренные факторы влияют, конечно, и на пропускную способность дискретных каналов, построенных на основе непрерывных.

Контрольные вопросы

1. Как осуществить сравнение пропускных способностей дискретного и непрерывного каналов?
2. Чему равно предельное значение нормированной пропускной способности дискретного канала при $h \to \infty$?

3. Когда пропускная способность дискретного канала может быть равна пропускной способности непрерывного канала?
4. Как влияют сосредоточенные помехи и замирания на пропускную способность каналов?

5.8. ИСПОЛЬЗОВАНИЕ ИНФОРМАЦИОННЫХ ХАРАКТЕРИСТИК В ИНФОРМАЦИОННО-ИЗМЕРИТЕЛЬНОЙ ТЕХНИКЕ

В информационно-измерительной технике, радиолокации, радионавигации и в других областях широко используют такие информационные характеристики, как энтропия, производительность и избыточность источника, скорость передачи информации и пропускная способность, коэффициенты согласования источников с каналами и др. Для примера рассмотрим особенности использования этих характеристик в информационно-измерительной технике.

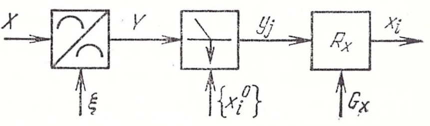


Рис. 5.5. Функциональная схема операций измерения

Информационно - измерительная техника (измерительная информационная техника) предназначена для получения опытным путем количественно определенной информации о разнообразных объектах материального мира [6]. Согласно ГОСТ 16263—70 измерение — это нахождение значения физической величины опытным путем с помощью специальных технических средств. В результате измерения получают численное отношение между измеряемой величиной и выбранной единицей измерения. Например, при измерении дальности в радиолокации в роли измеряемой величины выступает расстояние от радиолокационной станции до объекта, а в роли единицы измерения — метр. Обычно измеряемые величины рассматривают как случайные.

На рис. 5.5 показана типовая функциональная схема измерения [6]. Измеряемая величина X при наличии мешающего воздействия $\xi(t)$ (помехи) преобразуется в первом блоке в Y. Значения Y сравниваются в блоке сравнения с образцовыми мерами $\{x^0{}_i\}$. В результате сравнения определяют значение выходной величины y_j. В блоке принятия решения с использованием критерия G_x принимается статистическое решение R_x о том, что значение X является x_i. Такой подход к измерению обусловлен тем, что по y_j нельзя однозначно определить x_i. Можно лишь построить совокупность гипотез относительно того, каким является x_i на самом деле.

Неопределенность результата измерений определяют энтропией и рассчитывают в дискретном случае по формуле (1.8), а в непрерывном — по формуле (5.55). По значениям энтропии до и после измерения нетрудно определить количество информации $I(X, Y)$, полученной в результате измерения. Значение $I(Y, X)$ в дискретном случае рассчитывают по формулам (5.24),

(5.25), в непрерывном — по формулам (5.62), (5.63). Например, если X — непрерывная величина, а Y — дискретная, для определения количества информации используют формулу

$$I(Y,\ X)=\sum_{j}\int_{X} p(X,\ y_j)\log\frac{p(X,\ y_j)}{f(X)p(y_j)}\,dx.$$

Избыточность при измерениях обусловлена наличием вероятностных связей между результатами измерений. Наличие связей приводит к увеличению числа измерений n по сравнению с числом n_0 независимых измерений. *Удельное количество информации* на один отсчет определяют следующим образом:

$$I_1=I/n,\quad I_{1\text{макс}}=I/n_0,$$

где I — количество информации, полученное при измерениях. *Избыточность измерений*

$$r=(I_{1\text{макс}}-I_1)/I_{1\text{макс}}=1-n_0/n. \qquad (5.92)$$

Избыточность играет двоякую роль. С одной стороны, она вызывает появление лишних операций, уменьшение быстродействия, увеличение емкости устройств памяти и т. п., с другой,— она повышает помехоустойчивость и точность измерений.

Динамические свойства информационно-измерительных систем характеризуют скорость получения информации и пропускная способность. Если в течение времени измерений T получено количество информации $I(Y,\ X)$, то среднюю скорость получения информации определяют по формуле (5.26). Пропускную способность информационно-измерительной системы определяют по формуле (5.27). Коэффициент использования пропускной способности системы рассчитывают с помощью (5.17).

Если X является непрерывным сигналом с ограниченным значением верхней частоты F и нормальной плотностью распределения, ξ — аддитивной погрешностью преобразования X в Y, имеющей характер белого шума, то пропускная способность определяется формулой Шеннона (5.79). Максимальное количество информации, полученное в результате измерений, определяют по формуле (5.74). При малых погрешностях измерений, когда отношение сигнал/шум гораздо больше единицы, максимальное количество измеренной информации

$$V_{1\text{макс}}\cong FT\log(\mathcal{P}_c/\mathcal{P}_\xi).$$

При больших погрешностях измерений

$$V_{2\text{макс}}\cong FT\log(e\mathcal{P}_c/\mathcal{P}_\xi).$$

Согласование характеристик различных функциональных узлов информационно-измерительных систем выполняют с помощью рассмотренных соотношений и их функциональных схем.

Контрольные вопросы

1. Нарисуйте и поясните функциональную схему операций измерения.
2. Укажите основные особенности использования информационных характеристик в информационно-измерительной технике.

5.9. ВЫВОДЫ

1. Основными информационными характеристиками источников дискретных сообщений являются энтропия, условная энтропия, производительность и избыточность. Энтропия — это среднее количество информации, приходящееся на один символ сообщения. Если символы сообщения коррелированы, то энтропия уменьшается. Чтобы это учесть, вводят понятие условной энтропии. Она характеризует среднее количество информации, которое несет последующий символ сообщения при условии, что уже известен ряд предыдущих символов. Корреляционные связи и неравномерность распределения вероятностей появления символов характеризует избыточность источника — отношение количества информации, теряемой из-за воздействия этих факторов, к максимальному количеству информации, которое несет один символ.

2. Основными информационными характеристиками дискретных каналов являются скорость передачи информации, пропускная способность и коэффициент использования. Если помехи в канале отсутствуют, скорость передачи информации равна производительности кодера. Пропускной способностью канала без шумов называют максимальную скорость передачи информации при фиксированных условиях передачи и приема сигналов. Степень согласования источника с каналом характеризуют отношением скорости передачи информации к пропускной способности канала, называемым коэффициентом использования канала.

Среднее количество принятой информации определяют как разность среднего количества переданной информации и среднего количества информации, потерянной в канале из-за помех. Скорость передачи информации и пропускную способность каналов с шумами определяют через среднее количество принятой в единицу времени информации. Пропускная способность дискретных каналов с шумами полностью определяется скоростью передачи кодовых символов, вероятностью появления ошибки при передаче одного символа и основанием кода. Влияние помех проявляется в том, что для обеспечения сколь угодно малой вероятности ошибки необходимо увеличивать избыточность кода.

3. Сущность оптимального эффективного кодирования заключается в том, что разрабатывают такие коды, которые позволяют перевести неравномерное распределение вероятностей коррелированных символов в равномерное распределение вероятностей появления независимых кодовых символов. Конечная цель оптимального кодирования — приблизить скорость передачи информации к пропускной способности канала. Наличие шумов в канале приводит к тому, что эффективность оптимального кодирования

падает. Кодирование сообщений дискретных источников с памятью выполняют в два этапа: на первом устраняют корреляцию символов первичного алфавита, на втором выравнивают распределение вероятностей появления независимых кодовых символов. Эффективное оптимальное кодирование полезно для тех реальных каналов, в которых отношение сигнал/шум значительно превышает единицу.

4. Основными информационными характеристиками источников непрерывных сообщений являются энтропия, условная энтропия, эпсилон-энтропия, эпсилон-производительность, избыточность и объем информации. Формулу для энтропии источника непрерывных сообщений получают из формулы для энтропии источника дискретных сообщений путем предельного перехода. Среди всех источников с ограниченной и одинаковой мощностью сигналов наибольшей энтропией обладает источник с гауссовскими сигналами. Энтропия гауссовского белого шума является максимальной в классе гауссовских непрерывных сигналов. Поэтому помеха типа «гауссовский белый шум» является наиболее тяжелой и ее часто используют для оценки качества связи в наихудшем случае. Отсюда же следует целесообразность применения сигналов-переносчиков, по свойствам близких к гауссовскому белому шуму. Понятие эпсилон-энтропии вводят потому, что невозможно передать непрерывный сигнал абсолютно точно. Эпсилон-энтропия характеризует то среднее количество информации в одном независимом отсчете непрерывного случайного процесса, которое необходимо для воспроизведения этого процесса с заданной среднеквадратической погрешностью. Максимального значения эпсилон-энтропия достигает для гауссовских процессов. Эпсилон-производительность определяют как произведение средней скорости выдачи отсчетов на эпсилон-энтропию одного отсчета. Максимальный объем информации, выданной источником за время существования сигнала, равен произведению максимальной эпсилон-производительности на длительность сигнала. Избыточность источника характеризуют формулой (5.75).

5. Основными информационными характеристиками непрерывных каналов, как и дискретных, являются скорость передачи информации, пропускная способность и коэффициент использования. Пропускная способность гауссова непрерывного канала определяется его полосой пропускания и отношением сигнал/шум в канале. Чем шире спектр сигнала при прочих равных условиях, тем ближе пропускная способность канала к предельному значению. Поэтому целесообразно применять широкополосные (шумоподобные) сигналы-переносчики.

6. Предельные возможности согласования источников с каналами определены в теоремах кодирования Шеннона. Сущность этих теорем в том, что существуют способы оптимального кодирования, при которых вероятности ошибки (или для непрерывных сообщений среднеквадратические отклонения принятых сигналов от переданных) будут сколь угодно малы, если производитель-

ность источника меньше пропускной способности канала. Под оптимальным кодированием непрерывных сообщений в непрерывные сигналы понимают аналоговую модуляцию (см. § 3.1). Эффективность использования дискретными и непрерывными источниками пропускной способности непрерывного канала характеризуют коэффициентами (5.83), (5.87).

7. Сравнение пропускных способностей дискретного и непрерывного каналов показывает, что пропускная способность дискретного канала всегда меньше или равна пропускной способности соответствующего непрерывного канала. Равенство имеет место только тогда, когда и отношение сигнал/шум, и основание кода стремятся к бесконечности. Если вместо флуктуационной помехи в канале действует сосредоточенная помеха, то пропускная способность канала увеличивается. Если имеют место искажения и замирания сигналов в канале, то отношение сигнал/шум падает, а это приводит к уменьшению пропускной способности каналов.

8. Энтропия и производительность источника, скорость передачи информации и пропускная способность канала, коэффициенты (5.83), (5.87) широко используют для согласования характеристик функциональных узлов различных информационных систем.

Глава 6

ПОМЕХОУСТОЙЧИВОСТЬ ПЕРЕДАЧИ ДИСКРЕТНЫХ СООБЩЕНИЙ

6.1. ОСОБЕННОСТИ ОПРЕДЕЛЕНИЯ ПОМЕХОУСТОЙЧИВОСТИ ПЕРЕДАЧИ ДИСКРЕТНЫХ СООБЩЕНИЙ

Помехоустойчивость передачи дискретных сообщений определяется совместным действием большого числа взаимосвязанных факторов: избыточностью сообщений, способом кодирования, свойствами сигналов-переносчиков, видом модуляции, характером искажений сигналов и помех в канале, нарушением синхронизации передаваемых и принимаемых сигналов, способами демодуляции и декодирования принимаемых сигналов и др. Как уже отмечалось в § 1.5, проанализировать помехоустойчивость систем с учетом всех этих факторов не удается из-за сложности решения задач в такой общей постановке. Поэтому оценку помехоустойчивости проводят по этапам и на каждом этапе определяют влияние того или иного фактора на помехоустойчивость. Сравнение полученных результатов позволяет сделать вывод о том, какие факторы являются определяющими (см. § 6.6). Если можно получить решение задачи в более общей постановке, то определяют совместное влияние двух или нескольких факторов. Таким образом, методом последовательных приближений к реальным условиям помехоустойчивость определяют все более точно. Решение задач анализа помехоустойчивости реальных систем и синтеза проектируемых систем, оптимальных по критериям верности передачи сообщений, является предметом многих научно-технических исследований.

При определении помехоустойчивости важную роль играет выбор критерия верности с учетом конкретных условий передачи сообщений и исходных данных. Особенности выбора критериев верности рассматриваются в § 6.2. Задача выбора критерия во многом упрощается благодаря тому, что многие из критериев являются равноценными: их использование приводит к одним и тем же конечным результатам и схемным решениям.

Наиболее изучены задачи определения помехоустойчивости передачи сообщений для случаев, когда задан способ передачи и известны характеристики канала. В этих случаях задача по существу сводится к оценке верности различных способов приема сигналов — способов обработки сигналов в приемных устройствах. Поэтому наиболее разработанной частью теории помехо-

устойчивости является теория оптимального приема сигналов в различных условиях и для различных исходных данных. При приеме сигналов различают три основных класса задач: обнаружение сигналов, различение сигналов и восстановление сигналов. При обнаружении сигналов необходимо установить, имеется на входе приемника сигнал и помеха или только помеха. Эта задача, например, типична для радиолокации и для систем связи с пассивной паузой. При различении сигналов необходимо определить, какой сигнал из множества известных сигналов имеется на входе приемника. Сложность решения задачи определяется уже не свойствами отдельных сигналов и помех, а различием сигналов при наличии помех и искажений. Задачи различения сигналов типичны для дискретных систем связи с активной паузой. Задача восстановления сигналов существенно сложнее первых двух, сущность ее заключается в том, что необходимо на основании анализа принятого сигнала получить сигнал, который в том или ином смысле наименее отличается от переданного. Статистические характеристики передаваемого сигнала не всегда известны.

Методы приема сигналов классифицируют по ряду признаков: наличие информации о форме и фазе сигнала, способы додетекторной и последетекторной обработки, виды используемых детекторов и т. п. Различают когерентный, некогерентный, взаимокорреляционный, корреляционный и другие методы приема. Основные особенности этих методов, а также получивших практическое применение неоптимальных методов приема рассмотрены в § 6.3—6.5.

Для иллюстрации общих особенностей задач приема сигналов рассмотрим простейшую задачу обнаружения сигнала. Предположим, что на фоне аддитивной помехи $\xi(t)$ требуется обнаружить сигнал $s(t)$, функциональная форма и все параметры которого полностью известны. Требуется установить, является ли входной сигнал приемника суммой сигнала и помехи или на вход приемника поступает одна помеха. Для анализа верности приема сигналов в этой задаче можно применять как критерий максимального правдоподобия, так и критерий полной вероятности ошибки (см. § 4.3, 6.2). Конечные результаты решения задачи — одни и те же. В инженерной практике для построения оптимальных схем обнаружения используют отношение сигнал/шум, которое однозначно связано с указанными критериями.

Обычно принятый сигнал $z(t)=s(t)+\xi(t)$ подвергают такой обработке, которая математически описывается линейным оператором (см. приложение). В результате на выходе схемы обработки получают две составляющие. Одна полностью определяется свойствами сигнала, а другая — свойствами помехи. Отношение этих составляющих можно рассматривать как отношение сигнал/шум после обработки. Решающая схема приемника использует выходной сигнал схемы обработки, поэтому целесообразно отыскивать такой оптимальный оператор обработки принятого сигнала, который позволяет получить наибольшее отношение сиг-

нал/шум на выходе схемы при фиксированном отношении сигнал/шум на ее входе. Как правило, применяют оператор интегрирования в виде

$$E_{zx}=\int_0^T z(t)\,x(t)\,dt=(zx), \qquad (6.1)$$

где T — длительность сигнала $s(t)$; $x(t)$ — неизвестный сигнал, который необходимо определить в результате решения поставленной оптимальной задачи; (zx) — сокращенное обозначение операции (6.1).

Подставив в (6.1) значение $z(t)$, получим

$$E_{zx}=(sx)+(\xi x)=E_{sx}+\zeta, \qquad (6.2)$$

где E_{sx} — постоянная детерминированная величина, отражающая полезный сигнал; ζ — помеха на выходе схемы. Найдем оптимальный сигнал $x(t)$, максимизирующий функционал

$$E_{sx}=\int_0^T s(t)\,x(t)\,dt. \qquad (6.3)$$

Выражение (6.3) является взаимной энергией (2.11) сигналов $s(t)$ и $x(t)$. В § 2.1 показано, что взаимная энергия максимальна, когда сигналы полностью когерентны. Следовательно, $x(t)=\kappa s(t)$, где выбор постоянной κ не играет никакой роли — при взятии отношения сигнал/шум она сокращается. Поэтому $x_{\text{опт}}(t)=s(t)$, а

$$\max_{x(t)} E_{sx}=\int_0^T s^2(t)\,dt=E, \qquad (6.4)$$

где E — энергия сигнала $s(t)$.

Чтобы найти отношение сигнал/шум на выходе схемы обработки, найдем энергию помехи ζ. Предположим, что математическое ожидание помехи $M[\xi(t)]=0$, а дисперсия $M[\xi^2(t)]=\sigma^2$, тогда $M[\zeta]=0$. При выводе соотношения для энергии удобно воспользоваться тем, что оценку интервала корреляции $\Delta\tau$ помехи ξ определяют по формуле

$$\Delta\tau=\frac{1}{\sigma^2}\int_0^T K(\tau)\,d\tau, \qquad (6.5)$$

где $K(\tau)$ — корреляционная функция помехи $\xi(t)$.

Для того чтобы использовать (6.5), найдем математическое ожидание квадрата помехи ζ в виде

$$M[\zeta^2]=M\left[\int_0^T s(t)\,\xi(t)\,dt\int_0^T s(t_1)\,\xi(t_1)\,dt_1\right]=E_1. \qquad (6.6)$$

Поменяв операции интегрирования и определения математического ожидания местами, получим

$$E_1=\int_0^T s(t)\,dt\int_0^T s(t_1)\,M[\xi(t)\,\xi(t_1)]\,dt_1.$$

Так как $M[\xi(t)\,\xi(t_1)]=K(\tau)$, где $\tau=t-t_1$, то

$$E_1=\int_0^T s(t)\,dt\int_0^T s(t_1)\,K(t-t_1)\,dt_1.$$

Если ширина спектра помехи больше ширины спектра сигнала, то за время, равное интервалу корреляции помехи, сигнал не успеет значительно измениться. Поэтому $s(t_1)\simeq s(t)$ и

$$E_1\approx\int_0^T s^2(t)\,dt\int_0^T K(t-t_1)\,dt_1=E\int_0^T K(t-t_1)\,dt_1. \qquad (6.7)$$

Интеграл в (6.7) в соответствии с (6.5) равен $\sigma^2\Delta\tau$, следовательно,

$$E_1\simeq E\sigma^2\Delta\tau. \qquad (6.8)$$

Искомое максимальное отношение сигнал/шум на выходе схемы обработки

$$h_2=E^2/E_1=E^2/E\sigma^2\Delta\tau=\mathscr{P}T/\sigma^2\Delta\tau=h_1T/\Delta\tau, \qquad (6.9)$$

где $h_1=\mathscr{P}/\sigma^2$ — отношение сигнал/шум на входе схемы обработки. Из (6.9) следует, что оптимальный выбор сигнала $x(t)$, максимизирующего (6.3), максимизирует и (6.9), так как после сокращения на E знаменатель от E не зависит.

Введем *критерий эффективности схемы обработки*

$$g=h_2/h_1=T/\Delta\tau. \qquad (6.10)$$

Эффективность обработки тем больше, чем больше длительность сигнала превышает интервал корреляции помехи. Этот вывод лежит в основе различных методов накопления сигналов при обработке.

Если аддитивная помеха является белым шумом, то, учитывая (2.36), получаем

$$\sigma^2\Delta\tau=\int_0^T K(\tau)\,d\tau=\frac{\Omega_0}{2}\int_0^T \delta(\tau)\,d\tau=\frac{\Omega_0}{2}. \qquad (6.11)$$

Подставляя результат (6.11) в (6.9), находим

$$h_2=2E/\Omega_0=2\mathscr{P}T/\Omega_0=2h_1. \qquad (6.12)$$

Учтем, что $\Omega_0=\sigma^2/\Delta F$, где ΔF — полоса частот сигнала, тогда

$$h_2=2\mathscr{P}T\Delta F/\sigma^2=2T\Delta Fh_1=\nu h_1, \qquad (6.13)$$

$$g=2T\Delta F=\nu. \qquad (6.14)$$

Следовательно, если аддитивная помеха является белым шумом, то эффективность обработки равна базе сигнала, т. е. отношение сигнал/шум на выходе схемы в ν раз превышает отношение сигнал/шум на входе.

Для оценки минимальной вероятности ошибки решающей схемы, входным сигналом которой служит выходной сигнал схемы обработки, можно пользоваться формулой (4.31) с учетом того, что $h=\sqrt{h_2/2}$, так как (6.12) является отношением мощностей сигнала и помехи. Энергия белого шума, прошедшего схему обработки, определяется с учетом (6.8):

$$E_1=E\Omega_0/2. \tag{6.15}$$

Структурная схема оптимального линейного обнаружителя сигнала $s(t)$ на фоне белого шума представлена на рис. 6.1. Она включает генератор опорного сигнала $s(t)$, перемножитель, интегратор, и решающее устройство с оптимальным порогом $E/2$. Кроме того, в состав схемы обычно входит устройство синхронизации, предназначенное для запуска всей схемы в момент $t=0$ прихода сигнала, подключения решающей схемы ко входу интегратора при $t=T$, возврата схемы в исходное состояние после принятия решения. Схема реализует алгоритм (6.1). Если сигнал E_{zs} на выходе интегратора превышает порог $\sqrt{E}/2$, то на выходе решающей схемы отмечается появление сигнала на входе схемы. (Алгоритм обработки рассматривается как последовательность преобразований принятого сигнала, направленных на получение ответа на вопрос: есть ли сигнал на входе приемника?) Аппаратурная реализация линейной обработки сигналов может быть выполнена по различным схемам (см. § 6.3—6.6).

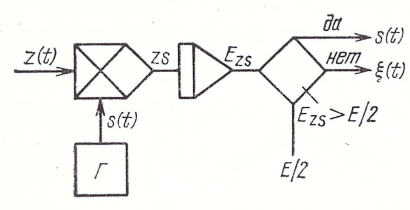


Рис. 6.1. Схема линейной обработки сигналов

В ГОСТ 21878—76 «Случайные процессы и динамические системы» оценку взаимной ковариационной функции двух эргодических сигналов $s(t)$ и $x(t)$ определяют по формуле

$$K^*_{sx}(\tau)=\frac{1}{T}\int_0^T s(t)\,x(t-\tau)\,dt.$$

Если сигналы центрированы, эта формула дает оценку взаимокорреляционной функции. Оценка взаимной энергии сигналов $E^*_{sx}=TK^*_{sx}(0)$, т. е. прямо пропорциональна взаимокорреляционной функции, коэффициентом пропорциональности служит длительность сигналов. В задачах обработки сигналов различные ковариационные и корреляционные функции сравнивают обычно при одинаковой длительности сигналов (см. § 6.3—6.6). Поэтому величину

$E^*_{sx}(0)$ рассматривают как существенную часть взаимной ковариационной или корреляционной функции, по которой принимают решение в приемнике, и считают что $E^*_{sx}(0)$ с точностью до постоянного множителя отражает корреляцию сигналов.

Операция (6.1) является общим видом *линейной обработки сигналов.* Частными случаями служат рассмотренное *когерентное обнаружение, когда* $x(t)=s(t)$, *интегральный прием*, когда $x(t)=1$; *взаимокорреляционный прием*, когда $x(t)=s(t-\tau)$; *корреляционный прием*, когда сигнал неизвестен и $x(t)=z(t-\tau)$; *прием на согласованный фильтр*, импульсная характеристика которого $x(t)=s(t-\tau)$. Фундаментальное значение операции (6.1) в том, что она используется при различных способах обработки сигналов. В § 6.3 операция (6.1) используется для оптимального различения полностью известных сигналов на фоне помехи (оптимальный когерентный прием), в § 6.4 — для оптимального различения сигналов с неопределенной фазой (некогерентный прием), в § 6.6 — для оптимального приема сигналов при неполной априорной информации.

Контрольные вопросы

1. В чем особенности определения помехоустойчивости передачи дискретных сообщений?
2. Какие задачи определения помехоустойчивости наиболее изучены?
3. Укажите сущность трех основных задач приема сигналов?
4. Как ставится задача обнаружения сигнала на фоне помехи?
5. Поясните особенности и количественные соотношения линейной обработки сигналов.

6.2. КРИТЕРИИ ВЕРНОСТИ ПЕРЕДАЧИ СООБЩЕНИЙ

При анализе помехоустойчивости применяют следующие критерии верности: среднего риска (байесов критерий), полной вероятности правильного приема (критерий идеального наблюдателя или критерий Котельникова), апостериорной вероятности правильного приема (критерий максимума апостериорной вероятности), минимаксный критерий, критерий Неймана — Пирсона, информационный критерий, критерий максимального правдоподобия и др. Рассмотрим сущность и особенности использования этих критериев и выберем подходящий для последующей оценки помехоустойчивости передачи дискретной информации.

6.2.1. Критерий среднего риска. Предположим, что передаются m сигналов: $s_i(t)$, $i=\overline{1, m}$. Через $z(t)=s_i(t)+\xi(t)$ обозначим принятый сигнал, где $\xi(t)$ — аддитивная помеха. Введем многомерную плотность $f(\mathbf{z}|s_i)$ распределения $\mathbf{z}$ при условии, что передавался s_i. Покажем, как получают аналитическое выражение для критерия среднего риска ρ.

Чтобы иметь возможность принять решение о том, какой сигнал пришел на вход приемника, все пространство $\mathbf{G}$ принимаемых сигналов необходимо определенным образом разбить на m подпро-

странств $G_i \in \mathbf{G}$, $i=\overline{1, m}$. Если принимаемый сигнал $z(t) \in G_i$, принимается решение, что передавался $s_i(t)$. Если на самом деле передавался сигнал $s_j(t)$, а сигнал $z(t)$ попал в подпространство G_i из-за действия помехи, то имеет место ошибка в принятии решения, а следовательно, и ошибка в передаче сообщений. Ошибки определяются свойствами передаваемых сигналов, характеристиками помех и выбором подпространств принятия решений.

На рис. 6.2 проиллюстрированы основные особенности принятия решения приемным устройством. Рассмотрен простейший случай передачи двух противоположных сигналов $s_1(t)$ и $s_2(t)$ с различными энергиями. Показано, как разбито пространство $\mathbf{G}$ на два подпространства G_1 и G_2 с границами γ_1 и γ_2. Размеры этих подпространств различны из-за того, что энергии передаваемых сигналов различны. Для определенности рассмотрена передача сигнала $s_1(t)$ при маломощной помехе $\xi_1(t)$ и более мощной $\xi_2(t)$. Когда помеха слабая, принятый сигнал $z_1(t)=s_1(t)+\xi_1(t)$ попадает в область G_1, $z_1(t) \in G_1$. Приемное устройство принимает правильное решение, что передавался сигнал $s_1(t)$. Когда помеха более мощная, принятый сигнал $z_2(t)=s_1(t)+\xi_2(t)$ попадает в область G_2, $z_2(t) \in G_2$. Приемное устройство принимает ошибочное решение, что передавался сигнал $s_2(t)$. Ошибка произошла из-за того, что под действием помехи $\xi_2(t)$ передаваемый сигнал $s_1(t)$ исказился и принятый сигнал $z_2(t)$ попал в область G_2. Рис. 6.2 наглядно показывает также влияние выбора сигналов $s_1(t)$, $s_2(t)$ и границ γ_1, γ_2 на правильность принятия решения. Например, если бы вместо границы γ_1 взять γ'_1, то и при слабой помехе произошла бы ошибка в принятии решения. Если вместо границы γ_2 выбрать ν'_2, то и при мощной помехе ошибки в принятии решения не было бы. Аналогично, если бы мощность сигнала $s_1(t)$ выбрать побольше, скажем, передавать $s'_1(t)$, то и при мощной помехе ошибки не произошло бы, так как $z'_2(t) \in G_1$.

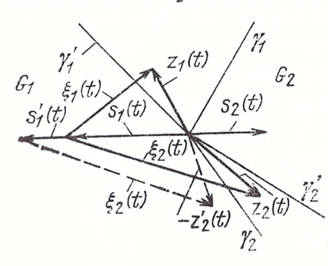


Рис. 6.2. Иллюстрация особенностей принятия решений приемником

Условная вероятность правильного приема сигнала

$$p(s_i \mid s_i) = \int\limits_{G_i} f(\mathbf{z} \mid s_i)\, d\mathbf{z}, \tag{6.16}$$

где интеграл является многомерным. Условная вероятность ошибки

$$p(s_j \mid s_i) = \int\limits_{G_j} f(\mathbf{z} \mid s_i)\, d\mathbf{z}. \tag{6.17}$$

Введем понятие потерь π_{ij}, которые возникают в случае принятия ошибочного решения, что был принят сигнал s_j, когда на самом

деле передавался s_i. (Естественно принять $\pi_{ii}=0$). Условный риск при передаче сигнала s_i

$$\rho_i=\sum_{j=1}^{m}\pi_{ij}p(s_j\mid s_i). \qquad (6.18)$$

Если $p(s_i)$ — априорная вероятность передачи сигнала $s_i(t)$, средний риск при передаче одного сигнала

$$\rho=\sum_{i=1}^{m}\rho_i p(s_i)=\sum_{i=1}^{m}\sum_{j=1}^{m}\pi_{ij}p(s_i)\,p(s_j\mid s_i). \qquad (6.19)$$

С учетом (6.17) получим

$$\rho=\sum_{i=1}^{m}\sum_{j=1}^{m}\pi_{ij}p(s_i)\int_{G_j}f(\mathbf{z}\mid s_i)\,d\mathbf{z}. \qquad (6.20)$$

Оптимальным по критерию среднего риска является тот способ передачи и приема сигналов, при котором минимизируется средний риск. Управляемыми переменными задач оптимизации являются характеристики сигналов, структура и параметры операторов преобразования сигналов в каналах, границы областей принятия решений. Исходными данными решения этих задач являются априорные вероятности $p(s_i)$, потери π_{ij}, характеристики помех или условные плотности вероятности $f(\mathbf{z}|s_i)$.

Критерий среднего риска является одним из наиболее общих. Частными случаями этого критерия являются критерии идеального наблюдателя, апостериорной вероятности правильного приема и др. Для использования критерия (6.20) требуется относительно большое количество таких исходных данных, которые на практике не всегда могут быть получены. Поэтому используют и другие критерии, лишенные этого недостатка.

6.2.2. Критерий идеального наблюдателя. Если в (6.20) принять $\pi_{ij}=1$ (все ошибки приводят к одинаковым последствиям), то (6.20) переходит в формулу *полной вероятности появления ошибки*

$$p=\sum_{i=1}^{m}\sum_{j\neq i}^{m}p(s_i)\int_{G_j}f(\mathbf{z}\mid s_i)\,d\mathbf{z}. \qquad (6.21)$$

Вероятность правильного приема

$$Q=1-p=\sum_{i=1}^{m}\int_{G_i}f(\mathbf{z}\mid s_i)\,p(s_i)\,d\mathbf{z}. \qquad (6.22)$$

Оптимальным по критерию идеального наблюдателя является тот способ передачи сообщений, при котором вероятность появления ошибки является минимальной, а вероятность правильного приема — максимальной. Управляемые переменные в задачах оптимизации те же, что и в п. 6.2.1. Упрощение решения задач дости-

гается благодаря тому, что не учитывается различие в последствиях ошибок. Максимальное значение Q достигается тогда, когда решение о том, что принятый сигнал относится к области G_i, принимается при выполнении условия

$$f(\mathbf{z}|s_i)p(s_i) > f(\mathbf{z}|s_j)p(s_j), \quad j \neq i. \tag{6.23}$$

Анализ $m-1$ условий (6.23) по существу равносилен построению алгоритма

$$\max_j [p(s_j) f(\mathbf{z}|s_j)] = p(s_i) f(\mathbf{z}|s_i). \tag{6.24}$$

Регистрируется тот сигнал $s_i(t)$, для которого априорная плотность распределения максимальна. Оптимальный приемник, работающий по этому алгоритму, В. А. Котельников назвал идеальным.

6.2.3. Апостериорная вероятность правильного приема. Так как $p(s_i)f(\mathbf{z}|s_i)=p(s_i|\mathbf{z})f(\mathbf{z})$, где $p(s_i|\mathbf{z})$ — апостериорная вероятность того, что передавался сигнал s_i при условии, что был принят $\mathbf{z}$; $f(\mathbf{z})$ — безусловная плотность распределения принятого сигнала, то в соответствии с формулой Байеса

$$p(s_i|\mathbf{z}) = p(s_i) f(\mathbf{z}|s_i) \Big/ \sum_{i=1}^{m} p(s_i) f(\mathbf{z}|s_i). \tag{6.25}$$

Если оптимальный приемник реализует алгоритм

$$\max_j [p(s_j|\mathbf{z})] = p(s_i|\mathbf{z}) \tag{6.26}$$

и принимает в этом случае решение о приходе сигнала $s_i(t)$, то говорят, что он обеспечивает *максимум апостериорной вероятности правильного приема.*

Сравнение алгоритмов (6.24) и (6.26) показывает, что они приводят к одному и тому же оптимальному решению. Следовательно, критерии идеального наблюдателя, максимума априорной плотности распределения и максимума апостериорной вероятности равносильны. Для решения задач оптимизации по этим критериям требуются одни и те же исходные данные.

6.2.4. Минимаксный критерий применяется в тех случаях, когда априорные вероятности появления сигналов в формуле (6.20) неизвестны. Задачу оптимизации решают для наихудшего распределения вероятностей $p(s_i)$, $i=\overline{1, m}$, и предполагают такое их распределение, которое доставляет максимум среднему риску. Используют критерий

$$\rho_1 = \min_{G_j} \max_{p(s_i)} \sum_{i=1}^{m} \sum_{j=1}^{m} \pi_{ij} p(s_i) \int_{G_j} f(\mathbf{z}|s_i)\, d\mathbf{z}. \tag{6.27}$$

Операция минимизации максимального риска, конечно, может выполняться путем выбора и других управляемых переменных. Минимаксный критерий не требует знания априорных вероятностей, но приводит к слишком осторожным решениям.

6.2.5. Критерий Неймана — Пирсона наибольшее распространение получил в радиолокации для решения задач обнаружения сигналов на фоне помех (см. § 6.1). Рассмотрим простейшую двуальтернативную ситуацию. Анализируя сигнал $z(t)$, приемник должен принять решение: есть сигнал, отраженный от объекта или нет его. Наличие сигнала обозначают символом 1, отсутствие — символом 0. Априорные вероятности $p(1)$ и $p(0)$ неизвестны. Могут быть ошибки двух родов: «*ложная тревога*» — принятие решения о том, что сигнал есть, когда его фактически нет, и «*пропуск цели*» — принятие решения о том, что сигнала нет, когда он фактически есть. Последствия этих двух ошибок далеко не равноценны.

Оптимальным по критерию Неймана — Пирсона считают такой способ обработки принятого сигнала, при котором обеспечивается *минимальная вероятность пропуска цели* $p(0|1)=\int_{G_0} f(\mathbf{z}|1)\,dz$ при заданной *вероятности ложной тревоги* $p(1|0)=\int_{G_1} f(\mathbf{z}|0)\,d\mathbf{z}$.

Приемник реализует алгоритм (6.26), но с учетом ограничения на вероятность ложной тревоги. Для двуальтернативной ситуации алгоритм (6.26) существенно упрощается и принимает вид $p(1|\mathbf{z})>$ $>p(0|\mathbf{z})$. Использовав (6.25), получим еще две равнозначные записи этого алгоритма:

$$p(1)f(\mathbf{z}|1)>p(0)f(\mathbf{z}|0)$$

и

$$f(\mathbf{z}|1)/f(\mathbf{z}|0)>p(0)/p(1).$$

Отношение в левой части последнего неравенства называют отношением правдоподобия и обозначают через Λ. Так как отношение в правой части неравенства неизвестно, его значение Λ_0 выбирают из условия обеспечения заданной вероятности ложной тревоги.

Перейдем от переменной $\mathbf{z}$ к Λ на основании равенств

$$f(\mathbf{z}|1)\,d\mathbf{z}=f(\Lambda|1)\,d\Lambda,\quad f(\mathbf{z}|0)\,d\mathbf{z}=f(\Lambda|0)\,d\Lambda.$$

Правые и левые части этих равенств выражают одну и ту же условную вероятность принять сигнал $\mathbf{z}$ при условии, что отраженный сигнал есть (первое равенство), и при условии, что его нет (второе равенство). При переходе от z к Λ пространство $\mathbf{G}$ преобразуется в числовую ось значений Λ, на которой значение Λ_0 представляет границу областей принятия соответствующих решений. Поэтому

$$p(0|1)=\int_{G_0} f(\mathbf{z}|1)\,d\mathbf{z}=\int_0^{\Lambda_0} f(\Lambda|1)\,d\Lambda,$$

$$p(1|0)=\int_{G_1} f(\mathbf{z}|0)\,d\mathbf{z}=\int_{\Lambda_0}^{\infty} f(\Lambda|0)\,d\Lambda.$$

Значение Λ_0 определяют из условия $\int\limits_{\Lambda_0}^{\infty} f(\Lambda|0)\,d\Lambda = \alpha_т$, где $\alpha_т$ — заданная вероятность ложной тревоги.

Следовательно, *оптимальный приемник Неймана — Пирсона* реализует алгоритм

$$\Lambda > \Lambda_0. \qquad (6.28)$$

Если условие (6.28) выполняется, принимается оптимальное решение о том, что сигнал есть.

6.2.6. Информационный критерий. В роли информационного критерия рассматривают отношение количества принятой информации $I(\mathbf{B}_2, \mathbf{B}_1)$ (5.24) к количеству переданной $H(\mathbf{B}_1)$:

$$\rho_2 = I(\mathbf{B}_2, \mathbf{B}_1)/H(\mathbf{B}_1) = 1 - H(\mathbf{B}_1|\mathbf{B}_2)/H(\mathbf{B}_1).$$

Оптимальным по этому критерию является такой способ передачи сообщений, при котором значение ρ_2 максимально. С учетом (5.37) можно заметить, что информационный критерий приводит к тем же результатам, что и критерии (6.22), (6.25), при решении многих задач.

6.2.7. Критерий максимального правдоподобия. *Функцией правдоподобия* называют плотность распределения $f(\mathbf{z}|s_i)$. Наиболее правдоподобна та гипотеза, для которой

$$f(\mathbf{z}|s_i) = \max_j [f(\mathbf{z}|s_j)]. \qquad (6.29)$$

Оптимальным является такой способ передачи сообщений, при котором реализуется алгоритм (6.29). Этот алгоритм является частным случаем алгоритма (6.24) при $p(s_i) = p(s_j) = 1/m$.

Критерий максимального правдоподобия получил наибольшее распространение потому, что он является относительно простым, не требует большого объема исходных данных, неплохо отражает большинство реальных условий передачи сообщений, аппаратурная реализация *алгоритмов максимального правдоподобия* относительно проста, результаты применения этих алгоритмов во многом совпадают с теми, которые получают с помощью многих других критериев. Поэтому в дальнейшем основным критерием верности служит критерий максимального правдоподобия.

Контрольные вопросы

1. Дайте сравнительную характеристику основных критериев верности.
2. Почему критерий среднего риска не получил широкого распространения?
3. В чем сущность и особенности применения критерия максимального правдоподобия?

6.3. ОПТИМАЛЬНЫЙ КОГЕРЕНТНЫЙ ПРИЕМ ДИСКРЕТНЫХ СИГНАЛОВ

Для *оптимального когерентного приема* необходимо выполнение следующих условий: передаваемые сигналы полностью известны и могут быть точно воспроизведены в приемном устройстве;

канал связи гауссов с постоянными параметрами, искажения сигналов в канале отсутствуют; спектральная плотность аддитивной помехи известна; синхронизация принимаемых и опорных сигналов является идеальной.

Постановка задачи оптимального когерентного приема дискретных сигналов следующая. Передаваемые сигналы $s_i(t)$, $i=\overline{1, m}$, поступают на вход гауссова канала с флуктуационной помехой $\xi(t)$, принимаемый сигнал

$$z(t)=s_i(t)+\xi(t). \tag{6.30}$$

Требуется построить оптимальный алгоритм различения сигналов приемником и оценить потенциальную помехоустойчивость оптимального когерентного приема.

6.3.1. Оптимальный алгоритм когерентного приема. Для построения оптимального алгоритма используем критерий максимального правдоподобия (6.29). Так как передаваемые детерминированные сигналы $s_i(t)$ в месте приема известны, то функция правдоподобия $f(\mathbf{z}|s_i)$ как плотность распределения принятого сигнала $z(t)$ полностью определяется n-мерной плотностью распределения помехи $f_1(\xi)$. Поэтому

$$f(\mathbf{z}|s_i)=f_1(\xi). \tag{6.31}$$

Воспользуемся (2.43), тогда

$$f(\mathbf{z}\,|\,s_i)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{2\sigma^2}\sum_{k=1}^{n}\xi^2_k\right), \tag{6.32}$$

где $\sigma^2=\Omega_0 F$; Ω_0 — спектральная плотность помехи; ΔF — полоса пропускания приемника. Выразим функцию правдоподобия (6.32) через разность принятого и передаваемого сигналов.

Значения ξ_k можно рассматривать как отсчеты помехи в ортогональном разложении Котельникова. В соответствии с равенством Парсеваля (2.14) энергия помехи, выделенная за время T действия сигнала,

$$\int_0^T \xi^2(t)\,dt=\frac{1}{2\Delta F}\sum_{k=1}^{n}\xi^2_k. \tag{6.33}$$

Отсюда следует, что

$$\sum_{k=1}^{n}\xi^2_k=2\Delta F\int_0^T \xi^2(t)\,dt,$$

$$\frac{1}{2\sigma^2}\sum_{k=1}^{n}\xi^2_k=\frac{2\Delta F}{2\sigma^2}\int_0^T \xi^2(t)\,dt=\frac{1}{\Omega_0}\int_0^T \xi^2(t)\,dt. \tag{6.34}$$

Поскольку $\xi(t)=z(t)-s_i(t)$, то

$$f(\mathbf{z}\,|\,s_i)=\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left\{-\frac{1}{\Omega_0}\int_0^T[z(t)-s_i(t)]^2\,dt\right\}. \qquad (6.35)$$

Следовательно, *алгоритм оптимального когерентного приема* (оптимального различения) сигналов

$$\max_i f(\mathbf{z}\,|\,s_i)=\max_i\left\{\frac{1}{(2\pi\sigma^2)^{n/2}}\exp\left(-\frac{1}{\Omega_0}\int_0^T[z(t)-s_i(t)]^2\,dt\right)\right\}. \qquad (6.36)$$

Очевидно, что операция (6.36) имеет место тогда, когда обеспечивается

$$\min_i\left\{\int_0^T[z(t)-s_i(t)]^2\,dt\right\},\quad i=\overline{1,\ m}. \qquad (6.37)$$

Алгоритм (6.37) является *оптимальным алгоритмом работы приемника Котельникова.* Этот приемник обеспечивает минимальную полную вероятность ошибки, поэтому он является оптимальным и по критерию идеального наблюдателя. Так как помехоустойчивость такого приемника является максимальной (предельно достижимой), то говорят, что он обладает «*потенциальной помехоустойчивостью*». Все реальные приемники имеют помехоустойчивость ниже потенциальной.

Анализируя алгоритм (6.37), можно заметить, что идеальный приемник Котельникова определяет в гильбертовом пространстве сигналов расстояние между принятым сигналом и всеми сигналами из ансамбля передаваемых и принимает решение, что передавался тот сигнал, к которому принятый ближе всего. В основе различных структурных схем оптимального приемника обычно лежит рассмотренная в § 6.1 схема линейной обработки сигналов. Отличие схем состоит в том, что используется различная аппаратурная реализация схем линейной обработки.

6.3.2. Оптимальный корреляционный приемник. Рассмотрим построение структурной схемы оптимального корреляционного приемника двоичных сигналов, в котором вычисляют взаимокорреляционные функции принятых и передаваемых сигналов. Для двоичных сигналов алгоритм (6.37) приводит к неравенству

$$\int_0^T[z(t)-s_1(t)]^2\,dt<\int_0^T[z(t)-s_2(t)]^2\,dt. \qquad (6.38)$$

Если (6.38) справедливо, принимается оптимальное решение о том, что передавался сигнал $s_1(t)$, если нет, то — сигнал $s_2(t)$.

Преобразуем неравенство (6.38) к виду, более удобному для аппаратурной реализации без применения сумматоров и квадра-

торов. После возведения разностей в квадрат и сокращения подобных членов получим

$$\int_0^T z(t)\, s_1(t)\, dt - \int_0^T z(t)\, s_2(t)\, dt > \frac{1}{2}\left[\int_0^T s^2{}_1(t)\, dt - \int_0^T s^2{}_2(t)\, dt\right]. \quad (6.39)$$

В левой части неравенства (6.39) взаимные энергии (6.40) принятого и опорных сигналов пропорциональны взаимокорреляционным функциям

$$E_{1z} = \int_0^T z(t)\, s_1(t)\, dt, \quad E_{2z} = \int_0^T z(t)\, s_2(t)\, dt \quad (6.40)$$

принятого сигнала с передаваемыми. Правая часть неравенства равна половине разности энергий передаваемых сигналов $(E_1 - E_2)/2$. Если энергии сигналов одинаковы, то неравенство (6.39) принимает простой вид

$$E_{1z} > E_{2z}. \quad (6.41)$$

Следовательно, при оптимальном различении сигналов решающая схема выделяет сигнал более коррелированный с принятым. Функционалы (6.40) полностью совпадают с функционалом (6.1).

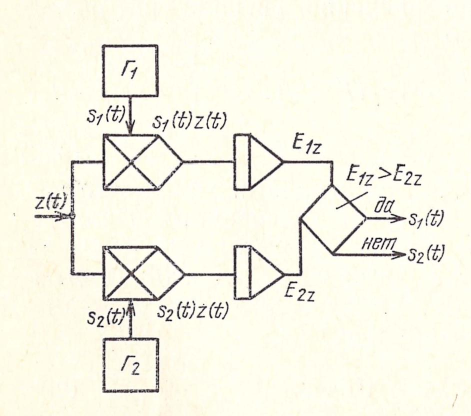


Рис. 6.3. Оптимальная структурная схема корреляционного приемника

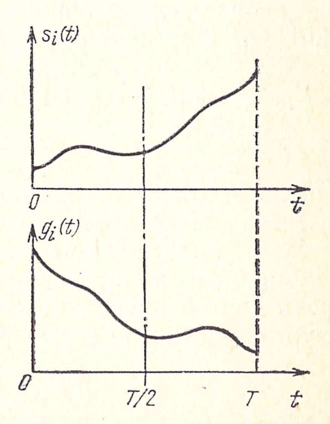


Рис. 6.4. Графики сигнала и характеристики согласованного фильтра

Поэтому схема корреляционного приемника включает две схемы линейной обработки сигналов, нагрузкой которых является решающая схема, реализующая условие (6.41). Структурная схема оптимального корреляционного приемника показана на рис. 6.3.

6.3.3. Оптимальный приемник на согласованных фильтрах. Недостатком схемы корреляционного приемника является необходимость использования опорных сигналов. Этот недостаток можно

устранить, если при построении структурной схемы приемника использовать согласованные фильтры.

Фильтр, импульсная характеристика которого

$$g_i(\tau)=s_i(T-\tau), \qquad (6.42)$$

называют согласованным с сигналом $s_i(t)$. Импульсной реакцией его $s_i(T-\tau)$ является сигнал $s_i(t)$ в обратной временной последовательности (начиная с момента времени $t=T$ и кончая $t=0$). График $g_i(t)$ (рис. 6.4) получается как зеркальное отражение графика $s_i(t)$ относительно вертикали, которая делит интервал $(0, T)$ пополам (штрих-пунктирная прямая на рис. 6.4). При $t=0$ $g_i(t)=0$, так как отклик фильтра не может предшествовать воздействию. При $t>T$ $g_i(t)$ может иметь произвольное значение, так как после взятия отсчета в момент времени $t=T$ выполняется сброс накопленного на фильтре значения сигнала, схема обработки возвращается в исходное состояние и готова к следующему циклу работы.

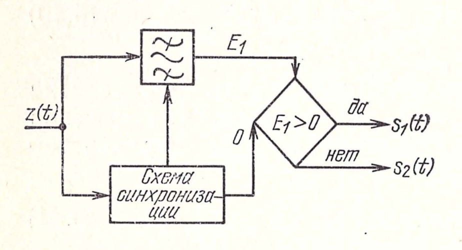


Рис. 6.5. Оптимальный приемник на согласованном фильтре

При воздействии сигнала $z(t)$ на вход согласованного фильтра значение сигнала на выходе фильтра в момент отсчета T равно

$$E_{iz}=\int_0^T z(t-\tau)\,g_i(\tau)\,d\tau=\int_0^T z(T-\tau)\,s_i(T-\tau)\,d\tau=\int_0^T z(t)\,s_i(t)\,dt. \qquad (6.43)$$

Сигналы (6.43) полностью совпадают с сигналами (6.40), что подтверждает возможность использования согласованных фильтров в схеме оптимального приемника. На первый взгляд кажется, что схема оптимального приемника двоичных сигналов должна включать два согласованных фильтра. Однако схему можно упростить, если неравенство (6.41) представить в виде

$$E_{\Delta s z}=\int_0^T z(t)\,[s_1(t)-s_2(t)]\,dt>0 \qquad (6.44)$$

и выбрать согласованный фильтр с характеристикой

$$g(t)=s_1(T-\tau)-s_2(T-\tau). \qquad (6.45)$$

На рис. 6.5 показана структурная схема оптимального приемника двоичных сигналов в одном согласованном фильтре. Устройство синхронизации обеспечивает запуск схемы в момент прихода сигнала ($t=0$) и отсчет значения сигнала E_1 на выходе фильтра в момент $t=T$. Линейная обработка сигналов согласованным фильтром существенно упрощает схему приемника. Задача синтеза оптимального приемника сводится к задаче синтеза оптимального линейного фильтра с постоянными параметрами. Для ее решения

можно применить разложение (6.45) по ортогональным **полиномам** и методы теории аппроксимации функций. Согласованный фильтр для сигналов произвольной формы можно построить и на основе неискажающей длинной линии.

6.3.4. Частотная и фазовая характеристики согласованного фильтра. Частотная характеристика

$$\mathcal{K}(i\omega)=\int_{-\infty}^{\infty} s(T-t)\,\mathrm{e}^{-i\omega t}dt=|\mathcal{K}(i\omega)|\,\mathrm{e}^{i\varphi_1(\omega)}. \tag{6.46}$$

Определим связь амплитудно-частотной характеристики фильтра $|K(i\omega)|$ со спектральной плотностью сигнала $\Omega(i\omega)$. Введем промежуточную переменную $\tau=T-t$, тогда

$$\mathcal{K}(i\omega)=\int_{-\infty}^{\infty} s(\tau)\,\mathrm{e}^{i\omega\tau}\mathrm{e}^{-i\omega T}dt=\Omega(-i\omega)\,\mathrm{e}^{-i\omega T}, \tag{6.47}$$

где $\Omega(-i\omega)$ — функция, комплексно-сопряженная со спектром сигнала

$$\Omega(i\omega)=\int_{-\infty}^{\infty} s(t)\,\mathrm{e}^{-i\omega t}dt=|\Omega(i\omega)|\,\mathrm{e}^{i\varphi(\omega)}. \tag{6.48}$$

Сравнение (6.47) и (6.48) показывает, что амплитудно-частотная характеристика согласованного фильтра

$$|\mathcal{K}(i\omega)|=|\Omega(i\omega)|, \tag{6.49}$$

а фазочастотная

$$\varphi_1(\omega)=-\varphi(\omega)-\omega T. \tag{6.50}$$

Выходной сигнал $s_1(t)$ фильтра при действии на его входе сигнала $s(t)$

$$s_1(t)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\Omega(i\omega)\,\Omega(-i\omega)\,\mathrm{e}^{i\omega(t-T)}d\omega=$$

$$=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\Omega(i\omega)|^2\,\mathrm{e}^{i\omega(t-T)}\,d\omega. \tag{6.51}$$

В соответствии с теоремой Релея при $t=T$ выходной сигнал

$$E_1=s_1(T)=\frac{1}{2\pi}\int_{-\infty}^{\infty}|\Omega(i\omega)|^2\,d\omega=\int_{-\infty}^{\infty}s^2(t)\,dt=E. \tag{6.52}$$

При $t=T$ все спектральные составляющие выходного сигнала совпадают по фазе и в сумме образуют максимальное значение $s_1(T)$, равное энергии входного сигнала. Во все другие моменты времени фазы спектральных составляющих различны $E_1<E$.

Эффективность согласованного фильтра как схемы линейной обработки сигналов была рассмотрена в § 6.1. Согласованная фильтрация может выполняться как для видеосигналов, так и для радиосигналов. Так как огибающая радиосигналов является «медленной» функцией по сравнению с высокочастотным колебанием, то допустимая неточность взятия отсчета для видеосигналов может быть больше. Если берутся отсчеты огибающей радиосигнала на выходе согласованного фильтра, то согласование фильтра с входным сигналом может быть выполнено с точностью до фазы. Это во многом упрощает схему приемника.

Для некоторых сигналов согласованные фильтры оказываются сложными в настройке и регулировке. Поэтому находят применение *квазиоптимальные фильтры*, которые согласованы с сигналами только по полосе пропускания. Например, для радиоимпульса прямоугольной формы длительностью T оптимальная полоса фильтра с прямоугольной частотной характеристикой $\Delta f=1{,}37/T$. Отношение сигнал/шум на выходе квазиоптимального фильтра на 15—20% меньше по сравнению с согласованным фильтром [2, 9].

Если помеха не является белым шумом, то для получения минимальной мощности помехи на выходе согласованного фильтра значения коэффициента передачи фильтра должны быть прямо пропорциональны спектральной плотности входного сигнала и обратно пропорциональны спектральной плотности $\Omega_2(i\omega)$ помехи. В этом случае

$$\mathscr{K}_1(i\omega)=\frac{1}{\Omega_2(i\omega)}\Omega(-i\omega)\,e^{-i\omega T}. \qquad (6.53)$$

Множитель $1/\Omega_2(i\omega)$ можно рассматривать как коэффициент передачи дополнительного линейного фильтра, который включен на входе согласованного и «выравнивает» распределение спектра помехи.

6.3.5. Потенциальная помехоустойчивость оптимального когерентного приемника. Для определения потенциальной помехоустойчивости необходимо определить эквивалентное отношение сигнал/шум (6.12) для различения сигналов и использовать соотношение (4.31). Ошибка в различении двоичных сигналов происходит, когда

$$z(t)=s_1(t)+\xi(t), \qquad (6.54)$$

а неравенство (6.39) не выполняется из-за действия помехи. Подставим (6.54) в (6.39) и учтем, что неравенство не выполняется. После алгебраических преобразований получим

$$\int_0^T \xi(t)\,[s_1(t)-s_2(t)]\,dt<-\frac{1}{2}\int_0^T [s_1(t)-s_2(t)]^2\,dt. \qquad (6.55)$$

Вероятность выполнения неравенства (6.55) является, очевидно, *минимальной вероятностью появления ошибки.*

Левая часть неравенства представляет помеху ζ на выходе согласованного фильтра, а правая — полезный сигнал, энергия кото-

рого равна половине энергии E_0 разности сигналов. Используя (6.4), (6.9) и (6.15), получаем отношение сигнал/шум на выходе фильтра

$$h_2 = 2E^2_0/E_0\Omega_0 4 = E_0/2\Omega_0. \qquad (6.56)$$

С учетом (4.31) вероятность появления ошибки

$$p_0 = P(\zeta < -E_0/2) = [1 - F_К(\sqrt{E_0/2\Omega_0})]/2, \qquad (6.57)$$

где $F_К(\cdot)$ — функция Крампа (2.88).

Из (6.57) следует, что минимальная вероятность ошибки при оптимальном когерентном приеме определяется энергией разности сигналов и спектральной плотностью помехи.

С учетом (2.100) представим отношение сигнал/шум в виде

$$h_2 = d^2T/2\Omega_0, \qquad (6.58)$$

где d — расстояние между сигналами $s_1(t)$ и $s_2(t)$ в гильбертовом пространстве. Анализ (6.58) показывает, что минимальная вероятность ошибки падает с ростом расстояния между сигналами и длительности сигналов и растет, если увеличивается спектральная плотность помехи. Следовательно, потенциальная помехоустойчивость зависит не только от энергии сигналов, но и от их взаимной энергии (2.11).

6.3.6. Потенциальная помехоустойчивость систем с фазовой телеграфией (ФТ), частотной телеграфной (ЧТ) и с амплитудной манипуляцией (АМн). Рассмотрим, как влияет выбор разности энергии сигналов E_0 на потенциальную помехоустойчивость различных систем. Выразим E_0 через энергию отдельных сигналов E_1, E_2 и нормированную взаимокорреляционную функцию ρ_{12} сигналов, тогда

$$E_0 = E_1 - 2E_{12} + E_2 = E_1 - 2\rho_{12}\sqrt{E_1E_2} + E_2, \qquad (6.59)$$

где

$$\rho^2_{12} = E^2_{12}/E_1E_2.$$

Для систем с активной паузой, к которым относятся системы с ФТ и ЧТ, сигналы имеют одинаковую энергию E, поэтому

$$h_2 = \frac{2E(1-\rho_{12})}{2\Omega_0} = h_1(1-\rho_{12}). \qquad (6.60)$$

Если сигналы одинаковы по форме, но противоположны, например

$$s_1(t) = -s_2(t), \text{ то } \rho_{12} = -1 \text{ и } h_2 = 2h_1. \qquad (6.61)$$

Такие сигналы используют в системе с ФТ, когда разность фаз сигналов равна π. Если сигналы ортогональны, что соответствует ЧТ, то $\rho_{12}=0$ и

$$h_2 = h_1. \qquad (6.62)$$

В системах с пассивной паузой, к которым относится система с амплитудной манипуляцией, $E_1=E$, $E_2=0$, поэтому $E_0=E$, а

$$h_2 = h_1/2. \qquad (6.63)$$

Условно можно полагать, что для таких систем $\rho_{12}=0{,}5$.

Сравнение потенциальной помехоустойчивости различных систем показывает, что наибольшую помехоустойчивость обеспечи-

вает ФТ. При одинаковой минимальной вероятности ошибки энергия сигналов ФТ может быть в четыре раза меньше энергии АМн сигналов и в два раза меньше энергии сигналов ЧТ.

6.3.7. Потенциальная помехоустойчивости систем с относительной фазовой модуляцией (ОФМ). Существенным недостатком систем фазовой телеграфии является возможность режима «обратной работы»—когда из-за случайного изменения фазы («случайного перескока фазы») опорного генератора даже при отсутствии помех появляются ошибки в различении сигналов. Это явление долгое время тормозило широкое практическое использование ФТ. В 1954 году Н. Т. Петрович предложил эффективный метод борьбы с этим явлением путем использования не абсолютной, а относительной фазовой модуляции (ОФМ), и системы фазовой модуляции и телеграфии начало широко внедряться.

Обычно на вход манипулятора ОФМ подают «двоичную» последовательность видеоимпульсов, а с выхода снимают высокочастотное (ВЧ) колебание со скачкообразными изменениями фазы в моменты времени, соответствующие границам видеоимпульсов. При подаче на вход манипулятора видеоимпульса с единичной амплитудой на выходе происходит скачок фазы несущего колебания, а при подаче видеоимпульса с нулевой амплитудой фаза ВЧ колебания не изменяется. В демодуляторе с ОФМ фаза принимаемого радиоимпульса отсчитывается не относительно фазы опорного колебания, а сравнивается с фазой предыдущего радиоимпульса. Таким образом, ОФМ отличается специальной перекодировкой символов в манипуляторе.

Перекодировка символов исходной двоичной последовательности сводится к тому, что при передаче символа «1» происходит изменение передаваемого сигнала, а при передаче символа «0» сигнал не изменяется. Например, если исходная последовательность 01101, то последовательность, полученная при перекодировке — 1001; первый символ исходной последовательности является опорным. Для перекодировки необходима задержка выходной последовательности на время длительности одного сигнала. На выходе приемника ОФМ символ «1» регистрируется при совпадении полярностей двух соседних посылок, а символ «0» — если полярности противоположны.

При ОФМ ошибки при перескоке фазы опорного сигнала возникают только в одном символе. Последующие символы регистрируются правильно и режим «обратной работы» отсутствует. Ошибки при различении сигналов ОФМ возможны в результате появления одного из двух несовместных событий: знак данного элемента принят ошибочно, а знак предыдущего правильно; знак данного элемента принят верно, а предыдущего — ошибочно. Вероятность каждого из этих событий определяется как $p_0(1-p_0)$, а вероятность ошибки при ОФМ как вероятность появления того или другого события, поэтому

$$p_{01}=2p_0(1-p_0). \qquad (6.64)$$

Следовательно, режим обратной работы устраняется путем увеличения (примерно вдвое) вероятности появления ошибки. Таким образом, при ОФМ вероятность ошибки

$$p_{01} \approx 2p_0 = 1 - F\,(\sqrt{2h_1}). \qquad (6.65)$$

6.3.8. Потенциальная помехоустойчивость многопозиционных систем. При различении m сигналов для получения вероятности ошибки p_m необходимо анализировать условия совместного выполнения m—1 неравенства типа (6.38). Для симметричного канала с аддитивной помехой вероятность совместного выполнения этих неравенств

$$1 - p_m = (1 - p_0)^{m-1}, \qquad (6.66)$$

где p_0 — вероятность ошибки в системе с активной паузой и противоположными сигналами. Так как $p_0 \ll 1$, используя формулу бинома Ньютона, получим

$$p_m = (m - 1)\,p_0. \qquad (6.67)$$

Из формулы (6.67) следует, что при одинаковых энергиях сигналов в системах с разными m вероятность ошибки растет линейно с ростом m. Однако отсюда не следует, что m-ичные системы обладают меньшей потенциальной помехоустойчивостью, чем двоичные. Необходимо учесть, что один m-ичный равновероятный символ несет в $\log_2 m$ раз большее количество информации. Поэтому сравнение m-ичных и двоичных систем необходимо производить при одинаковой скорости передачи информации и при одинаковых энергиях сигналов.

Энергия сигналов m-ичной системы $E_m = E \log_2 m$, где E — энергия двоичных сигналов. Вероятность ошибки при различении m сигналов приближенно определяется соотношением

$$p_m \approx \frac{1}{2}\,(m - 1)\left[1 - F_{\text{к}}\left(\sqrt{\frac{m}{m-1}\,h_1 \log m}\right)\right]. \qquad (6.68)$$

Анализ выражения для вероятности ошибки удобнее производить, используя асимптотическую формулу для функции Крампа [1], справедливую при больших значениях x

$$1 - F_{\text{К}}(x) = \frac{2}{\sqrt{2\pi x^2}} \exp\left(-\frac{x^2}{2}\right). \qquad (6.69)$$

Тогда

$$p_m \approx \frac{(m-1)^2}{\sqrt{2\pi h_1 m \log m}} \exp\left(-\frac{m}{2\,(m-1)}\,h_1 \log m\right). \qquad (6.70)$$

Из формулы (6.70) следует, что при $m \to \infty$, $p_m \to 0$ $h_1 = \text{const}$. Это соответствует выводу, следующему из основной теоремы Шеннона о существовании оптимального кодирования для канала с шумом. Из (6.68) как частный случай при $m=2$ следует (6.57).

Сравнение (6.68), (6.70) и (6.57) показывает, что m-ичные системы имеют более высокую потенциальную помехоустойчивость, однако техническая реализация многопозиционного кодирования сложнее. Поэтому многопозиционные сигналы еще не нашли широкого практического применения, а для повышения помехоустойчивости бинарных систем используют корректирующие коды.

Оптимальный когерентный прием сигналов имеет следующие особенности: при равновероятных сигналах работа оптимального приемника не зависит от интенсивности помех — приемник переходит в инвариантный режим; при флуктуационных помехах не требуется фильтрация входных сигналов; помехоустойчивость оптимального приемника не зависит от ширины полосы пропускания.

Контрольные вопросы

1. Как ставится задача оптимального различения дискретных сигналов?
2. Как определяется функция правдоподобия для когерентного различения сигналов?
3. Запишите и поясните алгоритм оптимального когерентного приема.
4. Нарисуйте и поясните работу схемы оптимального корреляционного приемника.
5. Какие амплитудно-частотные и фазочастотные характеристики имеет согласованный фильтр?
6. Нарисуйте и поясните схему оптимального приемника на согласованном фильтре.
7. Как определяют потенциальную помехоустойчивость оптимального приемника?
8. Сравните потенциальную помехоустойчивость систем с ФТ, ЧТ, АМн и ОФМ.
9. Чем определяется потенциальная помехоустойчивость многопозиционных систем?

6.4. ОПТИМАЛЬНЫЙ НЕКОГЕРЕНТНЫЙ ПРИЕМ ДИСКРЕТНЫХ СИГНАЛОВ

Предположение о том, что все параметры передаваемых сигналов полностью известны на приемной стороне, на практике не всегда выполняется. Из-за случайного изменения параметров каналов, под действием других случайных факторов параметры принятых сигналов также становятся случайными. Если законы распределения параметров сигналов могут быть получены путем измерений, то отыскивают среднее значение функции правдоподобия (6.35) и потом это значение используют для отыскания оптимального алгоритма приема сигналов. Когда фаза φ принимаемых сигналов является случайной величиной с известной плотностью распределения $f_2(\varphi)$ (см. § 4.1), то прием сигналов называют *некогерентным*. Среднее значение функции правдоподобия определяют как математическое ожидание функции случайного аргумента. С учетом (4.7) среднее значение функции правдоподобия

$$M\left[f(\mathbf{z} \mid s_i)\right] = \int\limits_{-\pi}^{\pi} \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{\Omega_0}\int\limits_0^T \left[z(\varphi,\ t) - s_i(t)\right]^2 dt\right\} f_2(\varphi)\, d\varphi. \tag{6.71}$$

6.4.1. Алгоритм и схема оптимального некогерентного приемника. Алгоритм оптимального некогерентного приема принимает вид

$$\max_k M\,[f\,(\mathbf{z}\,|\,s_k)] = M[f\,(\mathbf{z}\,|\,s_i)]. \qquad (6.72)$$

Если условие (6.72) выполняется, то приемник принимает оптимальное по критерию максимального правдоподобия решение, что был передан сигнал $s_i(t)$. После интегрирования по φ в (6.71) фаза при построении алгоритма уже отсутствует. Используя (6.71) и (6.72), можно показать, что *оптимальный некогерентный приемник* выделяет огибающую взаимокорреляционной функции

$$A_k = \left| \int_0^T \dot{z}\,(t)\,\dot{s}^*_k\,(t)\,dt \right|,$$

где $\dot{s}^*_k(t)$ — сигнал, комплексно-сопряженный с $\dot{s}_k(t)$, и решает, что был передан тот сигнал $\dot{s}_i(t)$, для которого A_k имеет наибольшее значение в момент отсчета. Следовательно, приемником реализуется алгоритм: если

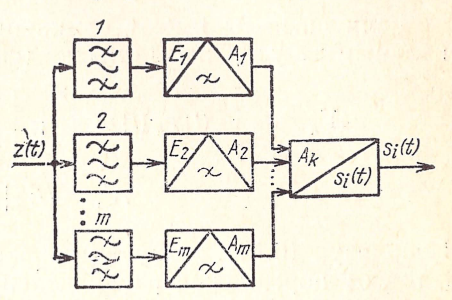


Рис. 6.6. Оптимальный некогерентный приемник

$$\max_k A_k = A_i, \quad k = \overline{1,\, m}, \qquad (6.73)$$

то принят сигнал $s_i(t)$.

Структурная схема оптимального некогерентного приемника приведена на рис. 6.6. Она содержит m согласованных фильтров, m детекторов и решающее устройство. На выходе k-го фильтра получается значение сигнала, пропорциональное взаимокорреляционной функции

$$E_k = \int_0^T \dot{z}\,(t)\,\dot{s}^*_k\,(t)\,dt. \qquad (6.74)$$

Амплитудный детектор выделяет огибающую A_k этого сигнала. В фиксированный момент времени отсчеты со всех детекторов поступают на решающую схему. Она принимает решение, что передавался сигнал $s_i(t)$, если соблюдается (6.73).

6.4.2. Вероятность ошибки для оптимального некогерентного приема. Выразим квадрат модуля взаимокорреляционной функции через принимаемые сигналы, передаваемые и сопряженные с ними по Гильберту сигналы, тогда

$$A^2_k = \left[\int_0^T z\,(t)\,s_k\,(t)\,dt \right]^2 + \left[\int_0^T z\,(t)\,\widehat{s}_k\,(t)\,dt \right]^2. \qquad (6.75)$$

Для определенности предположим, что передавался сигнал $s_1(t)$, тогда

$$A^2_k=\left[\int_0^T s_1(t)\,s_k(t)\,dt+\int_0^T \xi(t)\,s_k(t)\,dt\right]^2+$$

$$+\left[\int_0^T s_1(t)\,\widehat{s}_k(t)\,dt+\int_0^T \xi(t)\,\widehat{s}_k(t)\,dt\right]^2,\quad k=\overline{2,\,m}.$$

Если сигналы имеют одинаковую энергию и являются ортогональными в усиленном смысле (см. (2.72)), то

$$A^2_k=\left[\int_0^T \xi(t)\,s_k(t)\,dt\right]^2+\left[\int_0^T \xi(t)\,\widehat{s}_k(t)\,dt\right]^2=\xi^2_k+\zeta^2_k, \tag{6.76}$$

$$A^2_1=(E+\xi_1)^2+\zeta^2_1. \tag{6.77}$$

Используя (6.15), нетрудно показать, что случайные величины ξ_k и ζ имеют нормальное распределение с нулевым средним и дисперсией $E\Omega_0/2$. Огибающая A_k помехи распределена по закону Релея (2.78), который принимает следующий вид

$$f_k(A_k)=\frac{2A_k}{E\Omega_0}\exp\left(-\frac{A^2_k}{E\Omega_0}\right). \tag{6.78}$$

Случайная величина A_1 является «смесью» сигнала и помехи, поэтому она имеет обобщенное релеевское распределение (2.86), которое с учетом новых обозначений записывается в виде

$$f_1(A_1)=\frac{2A_1}{E\Omega_0}\exp\left(-\frac{A^2_1+E^2}{E\Omega_0}\right)I_0\left(\frac{2A_1}{\Omega_0}\right), \tag{6.79}$$

где $I_0(\cdot)$ — функция Бесселя (2.86).

Вероятность ошибки при оптимальном некогерентном приеме в данном случае является вероятностью того, что не будет выполняться неравенство

$$A_1>A_k,\quad k=\overline{2,\,m}. \tag{6.80}$$

Определим вероятность ошибки для простейшего случая $m=2$, когда

$$p_{22}=1-P(A_1>A_2)=P(A_2>A_1). \tag{6.81}$$

Так как A_1 и A_2 являются случайными величинами, то для определения вероятности ошибки необходимо зафиксировать значение A_1 и определить вероятность $p(A_1)$ того, что A_2 будет больше этого фиксированного значения, тогда

$$p(A_1)=\int_{A_1}^{\infty} f_2(A_2)\,dA_2. \tag{6.82}$$

После вычисления интеграла (6.82) необходимо учесть, что $p(A_1)$ является функцией случайного параметра A_1, и определить среднее значение

$$p_{22}=M[p(A_1)]=\int\limits_0^\infty p(A_1)f_1(A_1)\,dA_1. \qquad (6.83)$$

После интегрирования (6.82) получим

$$p_{22}=\int\limits_0^\infty \frac{2A_1}{E\Omega_0}\exp\left(-\frac{2E}{\Omega_0}-\frac{2A_1}{E\Omega_0}\right)I_0\left(\frac{2A_1}{\Omega_0}\right)dA_1. \qquad (6.84)$$

Используя табличное значение этого интеграла [1, 2], получаем

$$p_{22}=0{,}5\mathrm{e}^{-0{,}5h_1}, \qquad (6.85)$$

где $h_1=E/\Omega_0$.

Аналогично получают формулу для вероятности ошибки в многопозиционной системе при оптимальном некогерентном приеме

$$p_{2m}\cong 0{,}5\,(m-1)\,\mathrm{e}^{-0{,}5h_1}. \qquad (6.86)$$

Следовательно, как и при когерентном приеме,

$$p_{2m}\simeq(m-1)\,p_{22}. \qquad (6.87)$$

Анализируя (6.76), (6.77), (6.85) и (6.86), можно заметить, что для определения помехоустойчивости некогерентного приемника необходимо знать только энергию сигналов и спектральную плотность помехи. Сигнал «оставляет свои следы» в виде «деформации» распределения Релея в обобщенное распределение Релея.

На рис. 6.7 показаны графики зависимости вероятности ошибки в двоичной системе с активной паузой от отношения сигнал/шум на входе приемника при когерентном (кривая *1*) и некогерентном (кривая *2*) оптимальном приеме. Анализ графиков позволяет сделать вывод о том, что оптимальный некогерентный прием не столь существенно отличается по помехоустойчивости от оптимального когерентного приема. Все же при $h_1=5$ вероятность ошибки при некогерентном приеме $p_{22}\simeq 4{,}1\cdot 10^{-2}$, что примерно в 4 раза превышает вероятность ошибки при когерентном приеме. Увеличение вероятности ошибки является «платой» за неопределенность фазы сигнала. В то же время аппаратурная реализация некогерентного приема проще.

6.4.3. Помехоустойчивость оптимального некогерентного приемника двоичных сигналов с ЧМ. Особенности расчета помехоустойчивости некогерентного приема рассмотрим для приемника двоичных сигналов с ЧМ. Предположим, что канал является гауссовым с постоянными параметрами и флуктуационной помехой со спектральной плотностью $\Omega_0=10^{-8}$ мкВ/Гц, сигналы являются равно-

вероятными отрезками гармонических колебаний с прямоугольной огибающей — прямоугольные радиоимпульсы с частотой заполнения ω_1 и ω_2, длительностью $\tau=2$ мс, амплитуда сигналов на входе приемника $A=7$ мкВ, фильтры в субканалах приемника согласованы с сигналами. Рассчитаем вероятность ошибки приемника и сравним ее с вероятностью ошибки для когерентного приемника, т. е. реальную помехоустойчивость сравним с потенциальной.

На рис. 6.8 показана структурная схема некогерентного приемника двоичных ЧМ сигналов. Полосовые фильтры настроены на частоты ω_1 и ω_2, следующие за ними амплитудные детекторы выделяют огибающие A_1 и A_2. В решающей схеме отсчеты A_1 и A_2 сравниваются. Если $A_1>A_2$, схема принимает решение о том, что пришел сигнал $s_1(t)$, или что пришел сигнал $s_2(t)$.

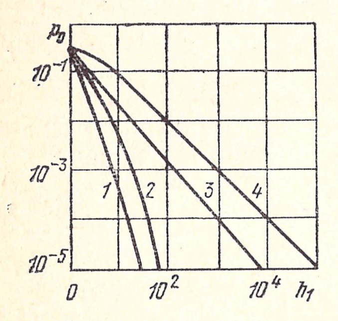


Рис. 6.7. Зависимости вероятности ошибки бинарной системы от отношения сигнал/шум

Рис. 6.8. Некогерентный приемник бинарной системы с ЧМ

Для расчета отношения сигнал/шум на входе приемника определим энергию импульса длительностью τ

$$E=P\tau=A^2\tau/2=49\cdot10^{-12}\cdot2\cdot10^{-3}/2=49\cdot10^{-15}\ \text{Вт}.$$

Вычислим отношение сигнал/шум

$$h_1=E/\Omega_0=49\cdot10^{-15}/10^{-8}\cdot10^{-6}=4{,}9.$$

Определим вероятность ошибки по формуле (6.80)

$$p_{22}=0{,}5e^{-h_1/2}=0{,}5e^{-2{,}45}\cong0{,}043.$$

В соответствии с (6.57) и (6.62) вероятность ошибки когерентного приемника $p_0=0{,}5\,[1-F_К(\sqrt{h_1})]\cong0{,}0139$. Следовательно, при некогерентном приеме вероятность ошибки возрастает примерно в 3,1 раза. Интересно сравнить полученные результаты с вероятностью ошибки при когерентном приеме многопозиционных сигналов. При $m=5$ расчеты по формуле (6.70) дают $p_{2m}\simeq0{,}003$. Реальная помехоустойчивость двоичного некогерентного приемника примерно в 13,8 раза хуже когерентного приемника при $m=5$. Как следует из (6.86), реальная помехоустойчивость двоичной системы всего лишь в четыре раза хуже реальной помехоустойчивости многопозицион-

ной системы с $m=5$. Поэтому некогерентный приемник системы с $m=5$ по помехоустойчивости близок к когерентному приемнику двоичных сигналов.

Таким образом, результаты анализа особенностей оптимального некогерентного приема показывают, что «платой» за отсутствие информации о фазе сигналов является снижение помехоустойчивости по сравнению с потенциальной. Наряду с этим упрощается схемная реализация некогерентного приемника.

Контрольные вопросы

1. В каких случаях решают задачу оптимального некогерентного приема?
2. В чем сущность оптимального алгоритма некогерентного приема?
3. Нарисуйте и поясните схему оптимального некогерентного приемника.
4. Как определяют вероятность ошибки для оптимального некогерентного приема?
5. Как связаны вероятности ошибки при некогерентном приеме двоичных и многопозиционных сигналов?
6. Сравните помехоустойчивость некогерентного и когерентного оптимальных приемников при выбранном отношении сигнал/шум.

6.5. НЕОПТИМАЛЬНЫЕ МЕТОДЫ ПРИЕМА ДИСКРЕТНЫХ СИГНАЛОВ

В реальных каналах связи из-за погрешностей формирования сигналов, рассинхронизации, искажений сигналов, недостаточной стабильности частоты, аппаратурных погрешностей функциональных узлов и ряда других причин условия оптимального приема часто нарушаются. Это приводит к снижению помехоустойчивости приема дискретных сигналов. Когда обеспечить условия оптимального приема трудно либо когда реализация оптимального приема связана с большим усложнением аппаратуры, на практике применяют неоптимальные методы приема. Рассмотрим особенности оценки помехоустойчивости неоптимального когерентного и некогерентного приема дискретных сигналов.

6.5.1. Неоптимальный когерентный прием. Предположим, что передаваемые сигналы $s_i(t)$, $i=\overline{1, m}$, равновероятны, ортогональны и имеют одинаковую энергию E. В результате нарушения условий оптимального приема, например из-за искажения сигналов в канале, на вход приемника поступает сигнал $z'(t)=s'_i(t)+\xi(t)$, где $s'_i(t)$ — искаженный сигнал $s_i(t)$. По отношению к сигналам $s'_i(t)$ приемник становится неоптимальным. Допустим, что энергия искаженных сигналов равна энергии передаваемых сигналов.

Неоптимальный когерентный приемник определяет взаимную энергию, пропорциональную взаимокорреляционной функции принятого и опорных сигналов,

$$E'_{kz}=\int_0^T z'(t)\, s_k(t)\, dt=E\rho'_{ik}+\zeta_k, \tag{6.88}$$

где

$$\rho'_{ik}=\frac{1}{E}\int_0^T s'_i(t)\,s_k(t)\,dt,$$

$$\zeta_k=\int_0^T s_k(t)\,\xi(t)\,dt, \qquad (6.89)$$

и принимает решение, что пришел тот сигнал $s_i(t)$, для которого величина E'_{kz} имеет наибольшее значение. Следовательно, *неоптимальный алгоритм когерентного приема* такой: если

$$\max_k E'_{kz}=E'_{iz},\quad k=\overline{1,\,m}, \qquad (6.90)$$

то передавался $s_i(t)$.

Если $\xi(t)$ — гауссовский белый шум со спектральной плотностью Ω_0, то E'_{kz} являются нормально распределенными случайными величинами с математическим ожиданием $M[E'_{kz}]=E\rho'_{ik}$ и дисперсией $E\Omega_0/2$. Для определенности предположим, что передавался сигнал $s_1(t)$, тогда

$$E'_{1z}=E\rho'_{11}+\zeta_1,\quad E'_{kz}=E\rho'_{1k}+\zeta_k,\quad k=\overline{2,\,m}.$$

Вероятность ошибки при неоптимальном когерентном приеме в данном случае является вероятностью того, что не будет выполнено неравенство

$$E'_{1z}>E'_{kz},\quad k=\overline{2,\,m}. \qquad (6.91)$$

Определим эту вероятность для простейшего случая $m=2$, тогда

$$p'_{20}=1-p(E'_{1z}>E'_{2z})=p(E'_{2z}>E'_{1z}).$$

Как и ранее (см. § 6.4), E'_{1z} и E'_{2z} являются случайными величинами, поэтому для определения вероятности ошибки необходимо вначале зафиксировать значение E'_{1z} и определить вероятность $p'(E'_{1z})$ того, что E'_{2z} будет больше этого фиксированного значения, тогда

$$p'(E'_{1z})=\int_{E'_{1z}}^{\infty} f_2(E'_{2z})\,dE'_{2z}. \qquad (6.92)$$

После вычисления интеграла (6.92) необходимо учесть, что $p'(E'_{1z})$ является функцией случайного параметра E'_{1z} и определить ее среднее значение

$$p'_{20}=M[p'(E'_{1z})]=\int_{-\infty}^{\infty} p'(E'_{1z})\,f_1(E'_{1z})\,dE'_{1z}. \qquad (6.93)$$

Учтем, что функции $f_1(E'_{1z})$ и $f_2(E'_{2z})$ являются нормально распределенными, тогда

$$p'_{20}=\int_{-\infty}^{\infty}\int_{E'_{1z}}^{\infty}\frac{1}{\pi E\Omega_0}\exp\left[-\frac{(E'_{2z}-E\rho'_{12})^2+(E'_{1z}-E\rho'_{11})^2}{E\Omega_0}\right]dE'_{1z}dE'_{2z}. \quad (6.94)$$

Этот интеграл вычисляется приближенно [9], оценка *вероятности появления ошибки при неоптимальном когерентном приеме двоичных сигналов*

$$p'_{20}\leqslant\{1-F_K[\sqrt{h_1}(\rho'_{11}-\rho'_{12})]\}/2. \quad (6.95)$$

Как частный случай при $\rho'_{12}=0$, $\rho'_{11}=1$ из (6.65) следует (6.57) с учетом (6.62).

Аналогично получают оценку для *вероятности ошибки при неоптимальном когерентном приеме m сигналов:*

$$p'_{m0}\leqslant(m-1)p'_{20}. \quad (6.96)$$

Погрешность оценки (6.96) не превышает 15% при $m\leqslant32$ и $p'_{20}\leqslant\leqslant10^{-4}$, что допустимо для инженерных расчетов.

Таким образом, помехоустойчивость неоптимального когерентного приема зависит от степени искажения сигналов в канале. Так как $\rho'_{ii}\leqslant1$, а $\rho'_{ik}\geqslant0$, то реальная помехоустойчивость, конечно, ниже потенциальной. С уменьшением ρ'_{ii} и увеличением ρ'_{ik} помехоустойчивость неоптимального когерентного приема падает. Для экспериментального определения ρ'_{ii}, ρ'_{ik} необходимо измерять искажения сигналов в реальном канале (см. § 4.4) и затем рассчитывать нормированные взаимокорреляционные функции. Из (6.95) следует, что для обеспечения требуемой помехоустойчивости при неоптимальном когерентном приеме требуется увеличивать отношение сигнал/шум, т. е. увеличивать энергию передаваемых сигналов.

6.5.2. Неоптимальный некогерентный прием. При неоптимальном некогерентном приеме определяется огибающая взаимокорреляционной функции и в соответствии с (6.73) приемник принимает решение, что пришел сигнал $s_i(t)$, если

$$\max_k A'_{kz}=A'_{iz}, \quad k=\overline{1, m}, \quad (6.97)$$

где A'_{kz} — огибающая, полученная с учетом того, что приходят искаженные сигналы. Из-за искажения сигналов, распределение огибающей даже на выходе тех субканалов, в которых нет полезного сигнала, является распределением «смеси» сигнала и шума, т. е. обобщенным релеевским распределением.

По-прежнему для определенности предположим, что передавался сигнал $s_1(t)$, тогда

$$A'^2_{1z}=(\rho'_{11}E+\xi_1)^2+\zeta^2_1, \quad (6.98)$$

$$A'^2_{kz}=(\rho'_{1k}E+\xi_k)^2+\zeta^2_k. \quad (6.99)$$

Если бы сигналы не были искажены, то $\rho'_{11}=1$, $\rho'_{ik}=0$ и выражения (6.98) и (6.99) перешли бы в (6.76) и (6.77).

Учтем, что случайные величины A'_{1z}, A'_{kz} имеют обобщенное релеевское распределение типа (6.79). Использовав это распределение и результаты п. 6.5.1, после приближенного вычисления интегралов типа (6.93), (6.94) получим *вероятности ошибки при неоптимальном некогерентном приеме*

$$p'_{22} \cong \frac{1}{2} e^{-h_1/2} I_0(\rho'_{12} h_1), \tag{6.100}$$

$$p'_{2m} \cong \frac{1}{2m} e^{-h_1/2} \sum_{i=1}^{m} \sum_{k \neq i}^{m} I_0(\rho'_{ik} h_1). \tag{6.101}$$

Когда $\rho'_{ik}=0$, формулы (6.100) и (6.101) переходят в формулы (6.85) и (6.86).

Анализируя (6.100) и (6.101), можно заметить, что помехоустойчивость неоптимального некогерентного приема ниже помехоустойчивости оптимального когерентного приема из-за того, что появляется корреляция между сигналами, обусловленная их искажениями. С ростом ρ'_{ik} растет вероятность появления ошибки неоптимального некогерентного приема. Например, вероятность ошибки неоптимального некогерентного приема двоичных сигналов в $I_0(\rho'_{12}h_1)$ раз больше этой вероятности для оптимального некогерентного приема.

6.5.3. Схемы неоптимальных некогерентных приемников. Практическое применение получили схемы узкополосного приема по огибающей, узкополосного приема по мгновенной частоте и широкополосного приема с последетекторным интегрированием. Рассмотрим особенности этих схем на примере систем с ЧМ.

Схема узкополосного приема по огибающей. В отличие от оптимальной схемы некогерентного приемника в этой схеме вместо согласованных фильтров стоят несогласованные «разделительные» фильтры. Такие фильтры проще реализуются, они согласуются с сигналами только по эффективной полосе пропускания Δf, их называют квазиоптимальными.

Гауссовский белый шум, воздействуя на квазиоптимальный линейный фильтр, сохраняет на выходе пассивного фильтра ($\mathcal{K}_0 \leqslant 1$) гауссовское распределение мгновенных значений и имеет мощность $P_\xi = \Omega_0 \Delta f$ (см. (4.43)). Амплитуда A_2 полезного сигнала на выходе фильтра определяется формой огибающей импульса, амплитудой A_1 и длительностью T входного сигнала, типом фильтра и его эффективной полосой пропускания Δf. Можно найти такое значение полосы $\Delta f_{\text{опт}}$, при котором отношение h_2 мощности сигнала к мощности помехи на выходе фильтра будет максимально в момент отсчета.

В качестве примера рассмотрим воздействие прямоугольного радиоимпульса на квазиоптимальный фильтр в виде одиночного колебательного контура с коэффициентом затухания $x_1 = 2\Delta f$. Ам-

плитуда выходного сигнала фильтра

$$A_2 = A_1 (1 - e^{-2\Delta f t}).$$

Отношение сигнал/шум в момент отсчета $t=T$

$$h_2 = \frac{A^2{}_1 (1 - e^{-2\Delta f T})^2}{2\Omega_0 \Delta f} = h_1 \frac{(1 - e^{-\Delta f T})^2}{\Delta f T}. \qquad (6.102)$$

Для определения $h_{2макс}$ продифференцируем (6.102) по Δf и приравняем результат нулю. Из полученного уравнения найдем

$$\Delta f_{опт} = 0,65/T. \qquad (6.103)$$

Подставив (6.103) в (6.102), получим

$$h_{2макс} \approx 0,82 h_1, \quad \vartheta_{макс} = 0,82,$$

$$\sqrt{\vartheta_{макс}} \approx 0,9,$$

где $\vartheta_{макс}$ — *эффективность квазиоптимальной фильтрации* (см. (6.10)). Сравнение (6.102) с результатом согласованной фильтрации (6.12) показывает, что эффективность квазиоптимальной фильтрации в данном случае примерно вдвое ниже согласованной.

Расчет оптимальных полос фильтров для других форм сигналов и других типов фильтров выполняют аналогично [9]. Эффективность квазиоптимальной фильтрации для различных пар «радиоимпульс — фильтр» приведена в табл. 3 [2]. Полученные оценки эффективности справедливы для случая приема одиночных импульсов, когда «перекрытием» соседних импульсов из-за переходных процессов в фильтрах можно пренебречь. При приеме случайной последовательности импульсов необходимо учитывать наличие остаточного напряжения на фильтре из-за переходных процессов, возникших при предыдущих посылках и переходных помехах, которые проявляются в том, что в момент отсчета принимаемый сигнал проходит и в фильтры, настроенные на частоты соседних позиций.

Таблица 3

Радиоимпульс	Тип фильтра	$\Delta f_{опт} T$	$\sqrt{\vartheta_{макс}}$
Прямоугольный	Идеальный П-образный	1,37	0,91
Прямоугольный	Гауссов	0,72	0,94
Гауссов	Идеальный П-образный	0,72	0,94
Гауссов	Гауссов	0,63	1,0
Прямоугольный	Одиночный резонансный контур	0,65	0,9

Расчеты показывают, что при приеме непрерывной последовательности импульсов $\Delta f'_{опт}$ примерно в два раза шире, чем оптимальная полоса $\Delta f_{опт}$ при приеме одиночных сигналов. Следовательно, *узкополосный прием по огибающей* проигрывает по мощ-

ности оптимальному некогерентному приему примерно в два раза, поэтому вероятность ошибки в данном случае

$$p'_{22} \approx 0{,}5e^{-h_1/4}. \tag{6.104}$$

Схема узкополосного приема двоичных ЧМ сигналов по мгновенной частоте состоит из узкополосного фильтра и частотного детектора. Последний настраивают так, чтобы при равенстве частоты настройки среднеарифметическому $(f_1+f_2)/2$ на выходе фильтра было нулевое напряжение (f_1 и f_2 — частоты сигналов $s_1(t)$ и $s_2(t)$). Решение принимается в зависимости от знака сигнала на выходе частотного детектора в момент отсчета. Если $\psi(f|s_i)$ — априорная условная плотность распределения мгновенной частоты f при передаче сигнала $s_i(t)$ и характеристика фильтра симметрична относительно f_0, то вероятность ошибки

$$p'_{22} = \int_{-\infty}^{0} \psi(f\,|\,s_i)\,df = \int_{0}^{\infty} \psi(f\,|\,s_i)\,df.$$

Плотность распределения $\psi(f|s_i)$ зависит от характеристик фильтра, девиации частоты сигнала $\Delta f_1=(f_2-f_1)/2$ и отношения сигнал/шум h_1.

Эффективная полоса фильтра в схеме с частотным детектором не менее чем в два раза больше $\Delta f_{\text{опт}}$ разделительных фильтров в схеме узкополосного приема по огибающей, потому что необходимо без искажений принимать сигналы $s_1(t)$ и $s_2(t)$. Отсюда следует, что помехоустойчивость узкополосного приема по огибающей и *узкополосного приема по мгновенной частоте с симметричным фильтром* примерно одинаковы. Вероятность ошибки оценивают по формуле (6.104).

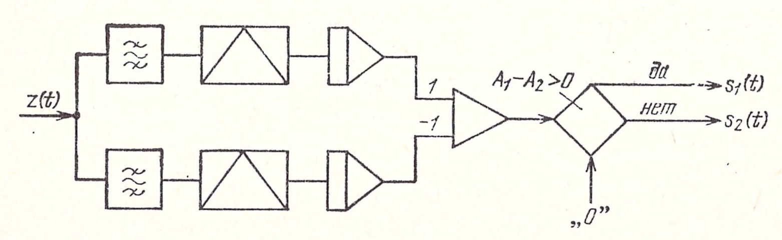


Рис. 6.9. Схема с последетекторным интегрированием сигналов до сумматора

Широкополосный прием с последетекторным интегрированием применяется при относительно низкой стабильности частоты. Вместо узкополосных разделительных фильтров используют широкополосные. При выборе полос пропускания фильтров учитывают возможные отклонения частоты сигнала из-за влияния дестабилизирующих факторов. Расширение полосы пропускания фильтра приводит к увеличению мощности шума на выходе фильтра и

к снижению помехоустойчивости приема. Для повышения помехоустойчивости применяют последетекторное интегрирование. Две эквивалентные по помехоустойчивости структурные схемы приема с последетекторным интегрированием сигналов приведены на рис. 6.9, 6.10. Схема рис. 6.10 является более простой, поэтому чаще применяется на практике.

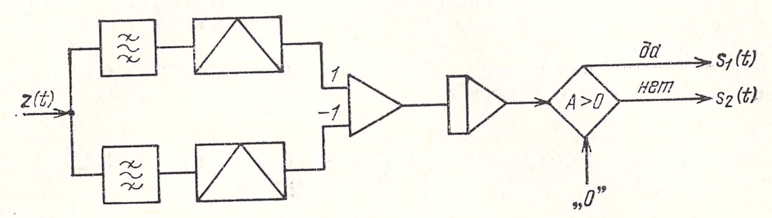


Рис. 6.10. Схема с последетекторным интегрированием сигналов после сумматора

Детектирование сигналов может быть квадратичным и линейным. Помехоустойчивость для обоих видов рассмотрена в [9]. При $\Delta fT \gg 1$ *вероятность ошибки при квадратичном детектировании*

$$p'_{02} \approx 0{,}5\left[1 - F_K\left(\sqrt{h_1\left(2 + \beta\Delta fT/h^2_1\right)^{-1}}\right)\right], \qquad (6.105)$$

где $1 \leqslant \beta \leqslant 2$. Для одиночного колебательного контура $\beta = 1$, для П-образного фильтра $\beta = 2$. Если $h_1 \gg \Delta fT$, то

$$p'_{02} \approx 0{,}5\left[1 - F_K\left(\sqrt{h_1/2}\right)\right]. \qquad (6.106)$$

Следовательно, при квадратичном детектировании имеется проигрыш по мощности сигнала в два раза по сравнению с оптимальным когерентным приемом.

При линейном детектировании и $h_1 \gg \Delta fT$ помехоустойчивость схемы с квадратичным детектором приближается к помехоустойчивости оптимального некогерентного приема. Вероятность ошибки в схеме с одиночным колебательным контуром в качестве фильтра

$$p''_{02} = 0{,}5\left[1 - F_K\left(1{,}39\sqrt{h_1} - 1{,}24\sqrt{\Delta fT}\right)\right], \qquad (6.107)$$

в схеме с П-образным фильтром

$$p'''_{02} \approx 0{,}5\left[1 - F_K\left(1{,}22\sqrt{h_1} - 1{,}1\sqrt{\Delta fT}\right)\right]. \qquad (6.108)$$

На практике вместо последетекторного интегратора часто применяют фильтр нижних частот, что приводит к зависимости выходного напряжения в момент отсчета от предыдущих посылок. Для наилучшего усреднения шумов и получения относительно малых остаточных напряжений от предыдущих посылок берут $\Delta fT = 1{,}5$—2,5. Энергетический проигрыш таких схем по сравнению с последетекторным интегрированием составляет 2—4 дБ.

Анализ конкретных схем неоптимального некогерентного приема дискретных сигналов показывает, что при $h_1 \gg \Delta fT$ эти схемы по помехоустойчивости приближаются к схемам оптимального некогерентного приема.

Контрольные вопросы

1. Из-за чего нарушаются условия оптимального когерентного приема в реальных каналах?
2. В чем сущность алгоритма неоптимального когерентного приема?
3. Как рассчитывают помехоустойчивость неоптимального когерентного приема?
4. В чем сущность алгоритма неоптимального некогерентного приема?
5. Как рассчитывают помехоустойчивость неоптимального некогерентного приема?
6. Сравните помехоустойчивость схем узкополосного приема по огибающей и по мгновенной частоте.
7. Нарисуйте схему и оцените помехоустойчивость широкополосного приема с последетекторным интегрированием.

6.6 ВЛИЯНИЕ ЗАМИРАНИЙ И РАССИНХРОНИЗАЦИИ СИГНАЛОВ НА ПОМЕХОУСТОЙЧИВОСТЬ

Рассмотрением влияния замираний и рассинхронизации сигналов на помехоустойчивость передачи сигналов завершается анализ влияния на помехоустойчивость тех основных факторов, которые вызывают существенное отклонение условий приема сигналов от идеальных. Полученные результаты, рассматриваемые совместно с результатами § 6.3—6.5, позволяют достаточно полно определить реальную помехоустойчивость различных систем.

6.6.1. Влияние замираний сигналов на помехоустойчивость. До сих пор рассматривались задачи приема сигналов для каналов с постоянными параметрами и аддитивной помехой. Рассмотрим, как полученные результаты обобщаются для каналов с переменными случайными параметрами, когда имеются замирания сигналов (см. § 4.1). В этом случае принятый сигнал $z(t)=\mathcal{K}(t)\times\times s(t)+\xi(t)$. Обозначим математическое ожидание случайного процесса $\mathcal{K}(t)$ через $\mathcal{K}_0$; предположим, что центрированный процесс $\mathcal{K}_1(t)=\mathcal{K}(t)-\mathcal{K}_0$ является стационарным случайным процессом с нулевым математическим ожиданием и дисперсией $\sigma^2_{\text{к}}$. Подставим $\mathcal{K}_1(t)$ в выражение для $z(t)$, тогда

$$z(t)=\mathcal{K}_0 s(t)+\mathcal{K}_1(t)s(t)+\xi(t)=\mathcal{K}_0 s(t)+\xi_1(t). \qquad (6.109)$$

Отличием рассматриваемого случая от ранее изученных является то, что сигнал умножается на среднее значение коэффициента передачи канала k_0, а эквивалентная аддитивная помеха $\xi_1(t)$ уже не является стационарным случайным процессом. Математическое ожидание помехи $M[\xi_1(t)]=0$, а дисперсия зависит от времени, так как

$$M[\xi^2_1(t)]=\sigma^2_{\text{к}}s^2(t)+M[\xi^2(t)].$$

Следовательно, помимо усреднения по множеству, необходимо применять усреднение и по времени.

Взаимная энергия принятого и передаваемого сигналов

$$E_{sz}=\int\limits_0^T z(t)\,s(t)\,dt=\mathcal{K}_0 E+\zeta+\int\limits_0^T \mathcal{K}_1(t)\,s^2(t)\,dt. \qquad (6.110)$$

Предположим, что замирания сигнала являются медленными и интервал корреляции процесса $\mathscr{K}_1(t)$ намного больше длительности сигнала, тогда

$$\int_0^T \mathscr{K}_1(t)\, s^2(t)\, dt \approx \mathscr{K}_1(t)\, E.$$

Энергия этого сигнала

$$M[\mathscr{K}^2_1(t) E^2] = E^2 M[\mathscr{K}^2_1(t)] = \sigma^2_{\text{к}} E^2. \qquad (6.111)$$

С учетом (6.8) отношение сигнал/помеха на входе решающей схемы

$$h_2 = \frac{\mathscr{K}^2_0 E^2}{\sigma^2_{\text{к}} E^2 + \sigma^2_{\xi} \Delta\tau E} = \frac{\mathscr{K}^2_0}{\sigma^2_{\text{к}} + \Delta\tau/T h_1} = \frac{\mathscr{K}^2_0 h_1}{\sigma^2_{\text{к}} h_1 + \Delta\tau/T}. \qquad (6.112)$$

Если $\xi(t)$ является гауссовским белым шумом, то $\Delta\tau \ll T$ и

$$h_2 = 1/U^2_0, \qquad (6.113)$$

где $U_0 = \sigma_{\text{к}}/\mathscr{K}_0$ — коэффициент вариации процесса $\mathscr{K}(t)$.

Из (6.113) следует, что отношение сигнал/шум полностью определяется вариацией коэффициента передачи канала. Для расчетов вероятностей ошибок необходимо учитывать, что из-за замираний сигнала отношение сигнал/шум изменяется пропорционально случайной величине $\mathscr{K}^2(t)$. Поэтому вероятности ошибок являются функцией случайного аргумента $\mathscr{K}$. Для расчета помехоустойчивости необходимо определять математические ожидания этих функций с учетом одномерного закона распределения коэффициента передачи канала. При релеевских замираниях коэффициент передачи подчиняется распределению (2.78) с учетом соотношений, приведенных в п. 4.1.3, поэтому математическое ожидание вероятности ошибки при оптимальном когерентном приеме двоичных сигналов

$$M[p_0(\mathscr{K})] = \int_0^\infty \frac{1}{2}\left\{1 - F_{\text{К}}\left[\frac{\mathscr{K}}{\mathscr{K}_0}\sqrt{h_1(1-\rho_{12})}\right]\right\} \frac{2\mathscr{K}}{\sigma^2_{\text{к}}} e^{-\mathscr{K}^2/\sigma^2_{\text{к}}} d\mathscr{K}.$$

Интегрирование по частям дает

$$M[p_0] = \frac{1}{2}\left[1 - \sqrt{\frac{h_1(1-\rho_{12})}{2 + h_1(1-\rho_{12})}}\right]. \qquad (6.114)$$

При $h_1 \gg 1$ можно применять приближенную формулу

$$M[p_0] = [2h_1(1-\rho_{12})]^{-1}. \qquad (6.115)$$

При релеевских замираниях (см. п. 4.1.3) и оптимальном некогерентном приеме двоичных сигналов в системах с активной паузой

$$M[p_{22}(\mathscr{K})] = \int_0^\infty \frac{1}{2} \exp\left(-\frac{\mathscr{K}^2}{\mathscr{K}_0}\frac{h_1}{2}\right) \frac{2\mathscr{K}}{\sigma^2_{\text{к}}} \exp\left(-\frac{\mathscr{K}^2}{\sigma^2_{\text{к}}}\right) d\mathscr{K} = \frac{1}{2 + h_1}. \qquad (6.116)$$

При квазирелеевских замираниях и оптимальном некогерентном приеме

$$M\,[p_{22}(\mathcal{K})]=\frac{\alpha^2+1}{h_1+2\alpha^2+2}\exp\left(-\frac{\alpha^2 h_1}{h_1+2\alpha^2+2}\right), \qquad (6.117)$$

где $\alpha=\mathcal{K}^2_0/(\sigma^2_к-\mathcal{K}^2_0)$. Формула (6.117) является наиболее общей для анализа влияния замираний сигналов на помехоустойчивость оптимального некогерентного приема. При $\sigma^2_к/\mathcal{K}^2_0\gg 2$ она переходит в (6.116), при $\sigma^2_к/\mathcal{K}^2_0\ll 2$ — в формулу средней вероятности ошибки при слабых (гауссовых) замираниях, когда коэффициент передачи распределен по нормальному закону с параметрами $\mathcal{K}_0$, $\sigma^2_к$.

Кривая *3* на рис. 6.7 показывает влияние релеевских замираний на среднюю вероятность ошибки (6.115) оптимального когерентного приема двоичных систем с активной паузой. Сравнение кривых *1*, *3*, показывает, что замирания сигналов существенно снижают помехоустойчивость. Графики, характеризующие помехоустойчивость оптимального когерентного приема при гауссовых замираниях и обобщенных релеевских, будут находиться в области между кривыми *1*, *3*, так как кривая *1* построена для канала без замираний.

Кривая *4* показывает влияние релеевских замираний на помехоустойчивость (6.116) оптимального некогерентного приема систем с активной паузой. При гауссовых и квазирелеевских замираниях кривые, характеризующие помехоустойчивость некогерентного приема, будут лежать в области между кривыми *2* и *4*, так как кривая *2* построена при отсутствии замираний. Сравнение графиков показывает, что замирания сигналов несколько больше влияют на помехоустойчивость оптимального некогерентного приема.

Используем исходные данные п. 6.4.3 и рассчитаем помехоустойчивость оптимального некогерентного приемника ЧМ сигналов при релеевских замираниях при условии, что $h_1=4{,}9$. В соответствии с формулой (6.116) $M[p_{22}]=1/(2+4{,}9)\simeq 0{,}145$. Следовательно, релеевские замирания приводят к тому, что помехоустойчивость рассматриваемого приемника ухудшается в 3,38 раза по сравнению со случаем без замираний и в 1,42 раза по сравнению с когерентным приемником при таких же замираниях.

Для борьбы с замираниями используют: автоматическую регулировку усиления (АРУ) сигналов и разнесенный прием сигналов, применяют широкополосные (шумоподобные) сигналы (см. § 3.3) и другие методы. Действие схемы АРУ сводится к тому, что она обеспечивает коэффициент усиления сигнала $\beta(t)=1/\mathcal{K}_1(t)$, тогда выходной сигнал $s_2(t)=\beta(t)\mathcal{K}_1(t)s_1(t)=s_1(t)$.

Наличие аддитивной помехи снижает эффективность, так как линейная обработка сигнала производится при флуктуирующей мощности белого шума.

Суть разнесенного приема заключается в том, что переданное сообщение в приемнике воспроизводится не по одному принятому

сигналу, а по двум или нескольким, несущим одно и то же сообщение. Обработка таких сигналов сводится к тому или иному виду суммирования с весом. В частности, это может быть определение среднего значения сигналов, выбор наибольшего из них и т. п. Различают методы разнесения по частоте, когда один и тот же сигнал одновременно излучается различными передатчиками на различных частотах, по времени — повторение передачи, пространственного разнесения, когда сигналы от одного и того же передатчика принимают на антенны, разнесенные в пространстве или с различной поляризацией.

6.6.2. Влияние рассинхронизации сигналов на помехоустойчивость. Помимо помех, большое влияние на помехоустойчивость оказывает нестабильность параметров приемника, например неточная работа схемы синхронизации, которая проявляется в «рассинхронизации» передаваемых и опорных сигналов. Из-за рассинхронизации вероятности ошибок когерентного и некогерентного приемов возрастают. Рассмотрим особенности оценки этих вероятностей при неидеальной временно́й синхронизации амплитудно-манипулированных сигналов. Для других видов модуляции эти вероятности определяют аналогично [14].

Задача ставится следующим образом: спектральная плотность Ω_0 белого шума $\xi(t)$, форма и параметры сигнала $s(t)$ известны точно, требуется определить, как влияет несовпадение момента t_1 прихода сигнала и момента t_2 генерирования опорного сигнала на вероятность ошибки приемника. Иначе говоря, необходимо определить влияние закона распределения $\omega(\tau)$, $\tau = t_2 - t_1$, на вероятность ошибки. В некогерентном приемнике τ характеризует погрешность времени взятия отсчета огибающей на выходе детектора (на входе решающей схемы).

При временно́м рассогласовании τ огибающей принимаемого и опорного сигналов условие принятия решения о приходе сигнала имеет вид

$$R_{sz}(\tau) = \frac{2}{\Omega_0} \int_0^T z(t)\, s(t - \tau)\, dt > z_0, \qquad (6.118)$$

где z_0 — порог сравнения в решающей схеме; нормировка с помощью множителя $2/\Omega_0$ выполнена для приведения $R_{sz}(\tau)$ к безразмерному виду.

Для определения вероятности ошибки как вероятности невыполнения условия (6.118) необходимо по известному распределению $\omega(\tau)$ найти распределение $\omega_1(u)$, где

$$u(\tau) = \frac{2}{\Omega_0} \int_0^T s(t)\, s(t - \tau)\, dt, \qquad (6.119)$$

а затем распределение $\omega_2(R_{sz})$ суммы

$$R_{sz}(\tau)=\frac{2}{\Omega_0}\zeta+u(\tau). \qquad (6.120)$$

По распределению $\omega_2(R)$ определяют математическое ожидание вероятности ошибки

$$M[p_э]=\int_0^{z_0}\omega_2(R_{sz})\,dR_{sz}. \qquad (6.121)$$

Если рассинхронизация обусловлена погрешностью работы схемы автоподстройки частоты приемника, то распределение $\omega(\tau)$ при больших отношениях сигнал/помеха на входе приемника можно полагать нормальным с нулевым средним и дисперсией σ_τ. Для прямоугольного видеоимпульса длительностью T и энергией E функция $u(\tau)$ имеет вид равнобедренного треугольника с основанием $2T$ и высотой $h_2=2E/\Omega_0$. Следовательно, $\tau\in[-T,\ T]$, $u(\tau)\in[0,\ h_2]$. Для удобства расчетов введем безразмерное среднеквадратическое значение времени расстройки $\sigma=\sigma_\tau/T$. После выполнения необходимых промежуточных преобразований получим

$$\omega_1(u)=\frac{c_0}{\sqrt{2\pi}\,\sigma}\exp\left[-\frac{(h_2-u)^2}{2h_2^2\sigma^2}\right], \qquad (6.122)$$

где c_0 — нормировочный множитель усеченного гауссовского распределения $\omega(\tau)$.

Распределение $\omega_2(R_{sz})$ определим как свертку [15] распределений $\omega_1(u)$ и $\omega_0(\zeta)$

$$\omega_2(R_{sz})=\int_{-\infty}^{\infty}\omega_1(x)\,\omega_0(R_{sz}-x)\,dx, \qquad (6.123)$$

что справедливо при условии независимости помех в цепи синхронизации и на входе приемника. Вычислив интеграл (6.123) с учетом того, что $\omega_0(\zeta)$ является гауссовским распределением с нулевым средним и дисперсией $h_2=2E/\Omega_0$, получим:

$$\omega_2(R)=\frac{[\Phi_Л(\alpha)+\Phi_Л(\beta)-1]\,c_0}{\sqrt{\pi h_2(1+h_2\sigma^2)/2}}\exp\left[-\frac{(R_{sz}-h_2)^2}{2h_2(1+h_2\sigma^2)}\right], \qquad (6.124)$$

где

$$\alpha(R_{sz})=\frac{1+\sigma^2R_{sz}}{\sigma\sqrt{1+h_2\sigma^2}};\quad \beta(R_{sz})=\frac{\sigma\,(h_2-\sigma R_{sz})}{\sqrt{1+h_2\sigma^2}};$$

$\Phi_Л(\cdot)$ — функция Лапласа (2.90).

В соответствии с (6.121) *математическое ожидание вероятности ошибки* при бинарных равновероятных сигналах

$$M[p_э]=\frac{1}{2}\left\{1-\Phi_Л\left(\frac{z_0}{2\sqrt{h_2}}\right)-c_0\Phi_Л\left[\frac{h_2-z_0}{\sqrt{h_2\,(1+h_2\sigma^2)}}\right]+\right.$$

$$+\frac{c_0}{\sqrt{2\pi h_2(1+h_2\sigma^2)}}\int\limits_0^{z_0}\exp\left[-\frac{(x-h_2)^2}{2h_2(1+h_2\sigma^2)}\right]\{\Phi_л[\alpha(x)]+\Phi_л[\beta(x)]\}\,dx\Bigr\}. \quad (6.125)$$

Если синхронизация идеальная, то $\sigma=0$, $z_0=h_1=E/\Omega_0$ и формула (6.125) переходит в известную (6.57) при h_2, определяемом из (6.62).

Анализ зависимостей типа (6.125) обычно производится численными методами с помощью ЦВМ. Численное интегрирование позволяет построить номограммы $M[p_з]=M[p_з(z_0, \sigma, h_2)]$.

Эти номограммы дают возможность определить влияние рассинхронизации на помехоустойчивость. На рис. 6.11 показана одна из таких номограмм. Графики зависимости вероятности ошибки от отношения сигнал/шум построены в логарифмическом масштабе при значении $z_0=h_1=E/\Omega_0$ и $\sigma=0$ (идеальная синхронизация),

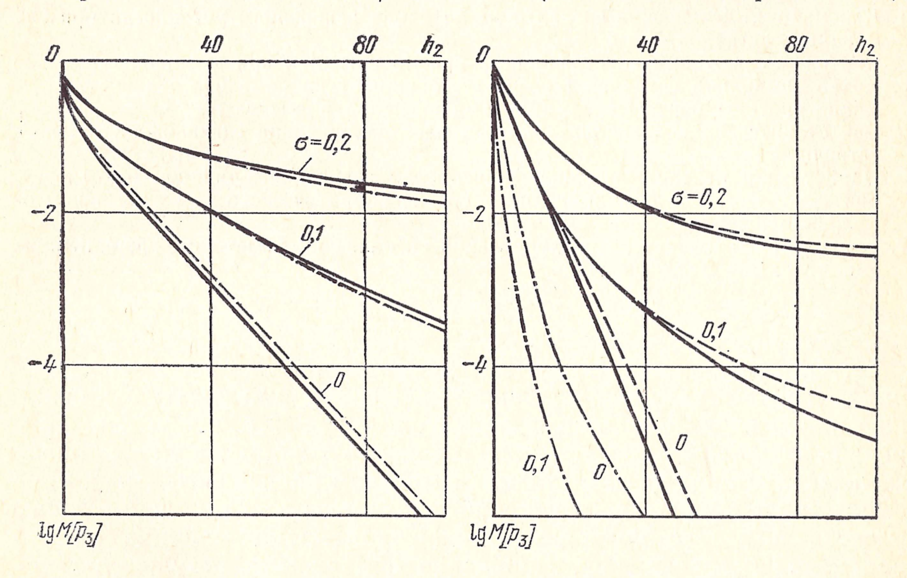


Рис. 6.11. Влияние рассинхронизации на помехоустойчивость приема АМн сигналов

Рис. 6.12. Влияние рассинхронизации на помехоустойчивость приема ЧМ и ФМ сигналов

$\sigma=0{,}1$ и 0,2. Помехоустойчивость приемника АМн сигналов существенно зависит от точности синхронизации. Например, при $h_2=60$ увеличение рассинхронизации на 20% проводит к увеличению вероятности ошибки более чем на два порядка. Влияние рассинхронизации и замираний сигналов на помехоустойчивость примерно одинаково. Для сравнения на рис. 6.11 показаны аналогичные кривые (штриховые линии) для оптимального некогерентного приема. Рассинхронизация примерно одинаково влияет на помехоустойчивость когерентного и некогерентного приемов.

На рис. 6.12 показано влияние рассинхронизации на помехоустойчивость приема ЧМ и ФМ сигналов. Кривые построены для когерентного приема ЧМ сигналов при $\sigma=0$; 0,1; 0,2, штриховые кривые отражают аналогичные зависимости для некогерентного приема. Штрих-пунктирные кривые показывают зависимости вероятностей ошибок для когерентного приема ФМ сигналов при $\sigma=0$; 0,1. Сравнение графиков рис. 6.11, 6.12 позволяет сделать следующие выводы. Рассинхронизация больше всего влияет на помехоустойчивость приема АМн сигналов. Влияние рассинхронизации на помехоустойчивость когерентного и некогерентного приемов примерно одинаковое, существенное различие проявляется лишь в области больших отношений сигнал/шум на входе решающего устройства.

Контрольные вопросы

1. Как находят отношение сигнал/шум на входе решающего устройства при замираниях сигналов?
2. Как определяют помехоустойчивость приема сигналов при замираниях?
3. Какие замирания сигналов больше всего влияют на помехоустойчивость?
4. Какие способы борьбы с замираниями сигналов применяют?
5. Как ставится задача анализа влияния рассинхронизации сигналов на помехоустойчивость?
6. Нарисуйте и поясните графики зависимости вероятности ошибки от отношения сигнал/шум и нормированного среднеквадратического значения времени рассинхронизации.
7. Сравните, как влияет рассинхронизация сигналов на помехоустойчивость приема АМн, ЧМ и ФМ сигналов.

6.7. ВЫВОДЫ

1. Определение помехоустойчивости передачи дискретных сигналов имеет следующие особенности. Необходимо учитывать большое число взаимосвязанных факторов, определяющих условия передачи сигналов. Из-за сложности решения задачи в целом оценку помехоустойчивости выполняют методом последовательных приближений от идеальных условий к реальным. На каждом этапе расчета помехоустойчивости обычно учитывают влияние одного определяющего фактора. Сравнение получаемых результатов позволяет выделить главные факторы и учесть совместное влияние основных из них. Наиболее изучены задачи оценки помехоустойчивости для случаев, когда способ передачи задан и характеристики канала известны. Эти задачи являются задачами теории оптимального приема.

Различают три основные задачи приема сигналов: обнаружение, различение и восстановление сигналов. В решении задач оптимального приема важную роль играют соотношения (6.4), (6.8) — (6.15), определяющие оптимальную линейную обработку (6.1) принятых сигналов, которая обеспечивает максимальное отношение сигнал/шум на выходе схемы. Оптимальная линейная обработка сигналов лежит в основе многих оптимальных способов приема.

2. Для определения помехоустойчивости передачи дискретных сообщений используют многие критерии верности: байесов критерий, критерий идеального наблюдателя, критерий максимального правдоподобия и др. Удобным для практики является применение критерия максимального правдоподобия, который не требует большого количества исходных данных. Решения задач во многих случаях совпадают с решениями, полученными с помощью других критериев.

3. Оптимальный когерентный прием дискретных сигналов на фоне помех целесообразен, когда форма и параметры передаваемых сигналов известны на приемной стороне абсолютно точно, искажения сигналов в гауссовом канале с постоянными параметрами отсутствуют, спектральная плотность помех известна, синхронизация принимаемых и опорных сигналов идеальная, аппаратурные погрешности отсутствуют. Оптимальный алгоритм когерентного приема сигналов (6.37) на фоне помех позволяет определить в гильбертовом пространстве сигналов расстояние между принятым сигналом и всеми передаваемыми сигналами и выбрать тот, к которому принятый ближе всего. Оптимальный приемник, работающий по этому алгоритму, обладает потенциальной помехоустойчивостью. Наибольшей потенциальной помехоустойчивостью обладают системы с ФТ, затем идут системы с ЧТ и АМн.

Идеальные условия когерентного приема на практике обычно не выполняются, поэтому основное значение этот прием имеет как эталонный.

4. Оптимальный некогерентный прием используют, когда фаза передаваемых сигналов неизвестна или случайна с известным законом распределения. Сущность алгоритма оптимального некогерентного приема (6.73) заключается в том, что берут отсчеты огибающих взаимокорреляционных функций принятого и передаваемых сигналов и принимают оптимальное решение, что был передан тот сигнал, для которого отсчет максимален. Некогерентный прием основан на том, что присутствие сигнала «деформирует» распределение Релея (распределение огибающей помехи) в обобщенное распределение Релея (распределение огибающей «смеси» сигнала и помехи).

5. На практике не всегда выполняются необходимые условия оптимального приема сигналов, поэтому нашли применение неоптимальные способы и схемы приема. При неоптимальном когерентном приеме реализуется алгоритм (6.90). Приемник принимает решение, что пришел тот сигнал, взаимокорреляционная функция которого с принятым в момент отсчета имеет максимальное значение. При неоптимальном некогерентном приеме алгоритм обработки сигналов (6.73) тот же, однако растет вероятность ошибки из-за искажений сигналов и применения неоптимальных схем.

Наибольшее распространение получили схемы узкополосного приема по огибающей, в которых вместо согласованных фильтров используются квазиоптимальные фильтры, узкополосного приема

двоичных ЧМ сигналов по мгновенной частоте, широкополосного приема с последетекторным интегрированием и др.

6. Большое влияние на помехоустойчивость оказывают замирания и «рассинхронизация» сигналов. Для анализа влияния замираний применяют усреднение по множеству вероятности ошибки с использованием одномерного закона распределения коэффициента передачи канала. Наибольшее влияние на помехоустойчивость оказывают релеевские замирания. Для борьбы с замираниями применяют автоматическую регулировку усиления сигналов, шумоподобные сигналы, разнесенный прием и другие способы.

Для анализа влияния рассинхронизации сигналов применяют усреднение по множеству вероятности ошибки с использованием одномерного закона распределения времени рассинхронизации. Рассинхронизация более всего снижает помехоустойчивость передачи АМн сигналов и менее всего — ЧМ и ФМ сигналов. Ухудшение помехоустойчивости из-за замираний и рассинхронизации сигналов примерно одинаково.

Глава 7

ПОМЕХОУСТОЙЧИВОСТЬ ПЕРЕДАЧИ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

7.1. ОСОБЕННОСТИ ОПРЕДЕЛЕНИЯ ПОМЕХОУСТОЙЧИВОСТИ ПЕРЕДАЧИ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

Структурная схема, предназначенная для иллюстрации основных особенностей определения помехоустойчивости передачи непрерывных сигналов, представлена на рис. 7.1. Модулирующий сигнал $u(t)$ и сигнал-переносчик $s_0(t)$ поступают на модулятор, в результате модуляции получается модулированный сигнал $s_1(t)$. Все последующие преобразования этого сигнала в канале обозначены одним оператором ψ. К этим преобразованиям относятся преобразования модулированных сигналов в многоканальных системах, линейные и нелинейные искажения модулированного сигнала и др. Передаваемый сигнал $s[u(t), t]$ поступает в линию связи (она представлена сумматором), в которой он взаимодействует с аддитивной помехой $\xi(t)$. Принимаемый сигнал

$$z(t)=s[u(t), t]+\xi(t). \tag{7.1}$$

Для восстановления полезного сигнала в приемнике — получения оценки $u^*(t)$ необходимо отфильтровать помеху с помощью фильтра и выполнить демодуляцию копии $s^*[u^*(t), t]$ модулированного сигнала.

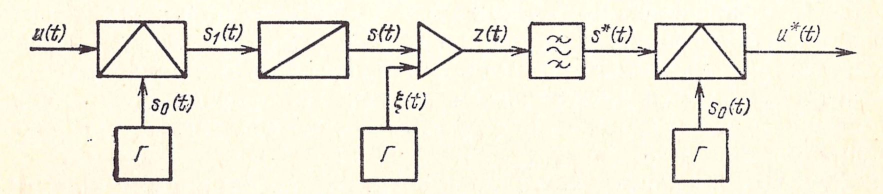


Рис. 7.1. Структурная схема системы передачи непрерывных сигналов

Из структурной схемы рис. 7.1 видно, что особенности определения помехоустойчивости передачи непрерывных сигналов включают все основные особенности определения помехоустойчивости передачи дискретных сигналов (см. § 6.1). Как и прежде, основная особенность заключается в том, что наиболее полно изучены задачи оценки верности способов приема сигналов, когда способы передачи и характеристики каналов известны.

Наряду с этим оценка помехоустойчивости передачи непрерывных сигналов имеет ряд принципиально новых особенностей. Во-первых, оценивается верность решения самой сложной задачи приема — восстановления (фильтрации) неизвестного полезного сигнала из искаженного модулированного сигнала на фоне помех. Следовательно, качество фильтрации во многом определяет верность. Во-вторых, основными полезными преобразованиями непрерывных сигналов в процессе передачи являются модуляция и демодуляция, поэтому по существу оценивается помехоустойчивость способа модуляции и модема. В-третьих, помимо модуляции и демодуляции, имеются и другие полезные и паразитные преобразования сигналов ψ, которые влияют на качество фильтрации. В четвертых, требуется уметь определять степень сходства $u(t)$ и $u^*(t)$, так как точное восстановление полезного сигнала невозможно. Следовательно, верность зависит от выбора метрики пространства сигналов, т. е. от того, как определяют расстояние между $u(t)$ и $u^*(t)$. В-пятых, верность существенно зависит от того, какое расстояние считается малым, т. е. от того, какие сигналы $u(t)$ и $u^*(t)$ можно считать близкими, например эпсилон-похожими (см. § 5.4). Наконец, в-шестых, выбор критерия верности существенно зависит от цели передачи сигналов, назначения системы и последующего использования принятых полезных сигналов $u^*(t)$.

Все эти особенности определения помехоустойчивости передачи непрерывных сообщений приводят к тому, что существенно усложняется решение задачи в общем виде. Поэтому конкретные задачи ставят так, чтобы использовать накопленный опыт и результаты, полученные при определении помехоустойчивости дискретных систем. Основным методическим приемом является ортогональное разложение полезного сигнала для того, чтобы задачу получения оценки $u^*(t)$ свести к задаче совместной оценки конечного числа случайных коэффициентов этого разложения. Этот прием лежит в основе определения потенциальной помехоустойчивости всех способов модуляции. Для оценки помехоустойчивости, как и ранее, удобно применять критерий максимального правдоподобия с последующим представлением оптимальных решений в виде, полезном для сравнительного анализа передачи непрерывных сообщений.

Рассмотрим постановку задачи оптимального приема. Если операторы M и ψ известны, то функция правдоподобия

$$f(z \mid s) = c \exp\left\{-\frac{1}{\Omega_0}\int_0^T \{z(t) - s[u(t), \ t]\}^2\, dt\right\}. \qquad (7.2)$$

Отыскание максимума (7.2) по всем $u(t)$ приводит к оптимальному алгоритму

$$\min_{u(t)} \{\varepsilon^2 [u(t)]\} = \min_{u(t)} \left[\frac{1}{\Omega_0}\int_0^T \{z(t) - s[u(t), \ t]\}^2\, dt\right]. \qquad (7.3)$$

В качестве оценки $u^*(t)$ полезного сигнала принимается та, которая доставляет минимум среднеквадратической погрешности. Как и в случае приема дискретных сигналов, задача сводится к определению расстояния в гильбертовом пространстве между принятым сигналом и всеми возможными передаваемыми сигналами $s[u(t),\ t]$ и выбору такого полезного сигнала, который соответствует минимальному расстоянию.

Непосредственно реализовать алгоритм (7.3) нельзя, так как неизвестны «возможные полезные сигналы». Рассмотрим, какие трудности это порождает. Раскрыв скобки в (7.3), получим

$$\varepsilon^2 = \frac{1}{\Omega_0}\left\{\int_0^T z^2(t)\,dt - 2\int_0^T z(t)\,s\,[u(t),\ t]\,dt + \int_0^T s^2\,[u(t),\ t]\,dt\right\}. \quad (7.4)$$

Первый интеграл в (7.4) характеризует энергию принятого сигнала, второй — взаимокорреляционную функцию переданного и принятого сигналов, третий — энергию передаваемого сигнала. Энергия принятого и передаваемого сигналов известны, неопределенность обусловлена тем, что сигнал $s[u(t),\ t]$ точно неизвестен и найти взаимокореляционную функцию, как это делалось раньше, невозможно.

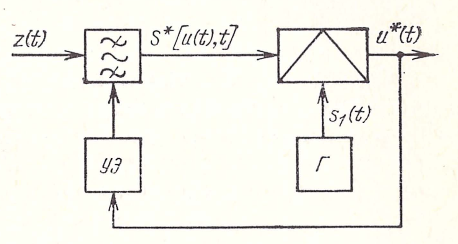


Рис. 7.2. Схема квазиоптимального приемника на согласованном фильтре с переменными параметрами

Поступают так. Располагая априорной информацией о способе модуляции и информационном параметре передаваемого сигнала, так обрабатывают принятый сигнал, чтобы дополучить недостающую часть информации об $s[u(t),\ t]$ и найти оценку $s^*[u(t),\ t]$. Эту оценку используют для оценки взаимокорреляционной функции получения $u^*(t)$. Такая идея лежит в основе работы *квазиоптимального следящего корреляционного приемника* и *квазиоптимального приемника на согласованном фильтре с переменными параметрами.*

Покажем, как работает квазиоптимальный приемник на согласованном фильтре с переменными параметрами (рис. 7.2). Структурная схема имеет основной информационный канала, который включает согласованный фильтр с переменными параметрами и демодулятор, на выходе которого получают оценку $u^*(t)$, а также канал обратной связи, который в зависимости от значения $u^*(t)$ с помощью управляющего элемента (УЭ) обеспечивает такое изменение параметров фильтра, чтобы в любой момент времени достигалось оптимальное согласование фильтра с непрерывно изменяемым сигналом $s^*[u(t),\ t]$. Вид демодулятора определяется способом модуляции. Следовательно, в таком приемнике в результате слежения за информационным параметром схема автомати-

ческого согласования фильтра обеспечивает получение текущих оптимальных оценок максимального правдоподобия $u^*(t)$.

Рассмотрение общей постановки оптимального приема показывает, что основной операцией является фильтрация — восстановление модулированного сигнала. Применяют *непрерывную* и *дискретную фильтрацию*. Рассмотрим сущность и основные особенности этих способов фильтрации.

7.2. ОПТИМАЛЬНАЯ ФИЛЬТРАЦИЯ НЕПРЕРЫВНЫХ СИГНАЛОВ

7.2.1. Оптимальная линейная фильтрация [13]. Покажем и сравним три основных подхода к оптимальной линейной фильтрации: определение частотных и импульсных характеристик оптимальных фильтров, восстановление сигнала как решение дифференциального уравнения, коэффициенты которого задаются статистикой сигналов. Такие задачи относят к задачам линейной фильтрации случайных сигналов на фоне помех. Впервые они были корректно поставлены и решены в 40-х годах советским математиком А. Н. Колмогоровым и американским ученым Н. Винером. Результаты решения этих задач легли в основу статистической теории линейной фильтрации.

Предположим, что $s[u(t), t]$ может рассматриваться как стационарный эргодический случайный процесс, корреляционная функция $K_1(\tau)$ или спектральная плотность $\Omega_1(\omega)$ которого известна. Канал связи будем считать гауссовым с аддитивной помехой $\xi(t)$, корреляционная функция $K_2(\tau)$ или спектральная плотность $\Omega_2(\omega)$ которой известна. Сигнал и помеха некоррелированны. Принятый сигнал описывается соотношением (7.1). Необходимо найти такой оптимальный линейный фильтр, который дает оценку $s^*[u(t), t]$, а в более общем случае оценку $u^*(t)$, в среднеквадратическом смысле наиболее близкую к переданному сигналу.

Обозначим оператор линейной фильтрации через L_1. Предположим, что допустимо только запаздывание сигналов в процессе фильтрации, тогда с учетом эргодичности процессов $s(t)$ и $\xi(t)$ среднеквадратическая погрешность примет вид

$$\varepsilon^2 = \langle\{L_1[z(t)] - s(t - t_0)\}^2\rangle, \tag{7.5}$$

где t_0 — время запаздывания сигнала. Преобразуя выражение (7.5) с использованием ранее сформулированных условий, получим

$$\varepsilon^2 = \langle\{L_1[s(t)] - s(t - t_0)\}^2\rangle + \langle\{L_1[\xi(t)]\}^2\rangle = \mathscr{P}_3 + \mathscr{P}_4, \tag{7.6}$$

где $\mathscr{P}_3$ — составляющая погрешности, обусловленная прохождением через фильтр сигнала; $\mathscr{P}_4$ — составляющая погрешности, обусловленная прохождением помехи.

Для минимизации мощности $\mathscr{P}_3$ и отсутствия искажений сигнала $s(t)$ фазочастотная характеристика искомого фильтра должна быть линейной:

$$\varphi_0(\omega) = \omega t_0, \tag{7.7}$$

а амплитудно-частотная характеристика $\mathscr{K}(\omega)$ должна выбираться из условия $\Omega_3(\omega)=[1-\mathscr{K}(\omega)]^2\Omega_1(\omega)$. Мощность $\mathscr{P}_4$ отклика фильтра на входной шум от фазовых соотношений не зависит, спектр помех $\Omega_4(\omega)=\mathscr{K}^2(\omega)\Omega_2(\omega)$. Суммарный спектр сигнала ошибки (7.6) на выходе фильтра

$$\Omega_\varepsilon(\omega)=\Omega_1(\omega)[1-\mathscr{K}(\omega)]^2+\Omega_2(\omega)\mathscr{K}^2(\omega). \qquad (7.8)$$

Оптимальную характеристику $\mathscr{K}_0(\omega)$ определим, дифференцируя (7.8) по $\mathscr{K}(\omega)$ и приравнивая производную нулю:

$$\mathscr{K}_0(\omega)=\Omega_1(\omega)/[\Omega_1(\omega)+\Omega_2(\omega)]. \qquad (7.9)$$

Подставив (7.9) в (7.8) получим, что

$$\Omega_\varepsilon(\omega)=\Omega_1(\omega)\Omega_2(\omega)/[\Omega_1(\omega)+\Omega_2(\omega)]. \qquad (7.10)$$

Минимальная среднеквадратическая погрешность

$$\varepsilon^2_{\text{мин}}=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\frac{\Omega_1(\omega)\Omega_2(\omega)}{\Omega_1(\omega)+\Omega_2(\omega)}d\omega. \qquad (7.11)$$

Выражение (7.11) дает минимально достижимую среднеквадратическую погрешность выделения оптимальным линейным фильтром сигнала из смеси с шумом. Никакой другой линейный фильтр не обеспечит лучшего в среднеквадратическом смысле отделения сигнала от помехи.

Коэффициент передачи *оптимального линейного фильтра*

$$\mathscr{K}_0(i\omega)=\frac{\Omega_1(\omega)}{\Omega_1(\omega)+\Omega_2(\omega)}\,e^{-i\omega t_0}. \qquad (7.12)$$

Этот фильтр называют оптимальным фильтром Колмогорова — Винера. Проанализируем характеристики оптимального фильтра. Для удобства анализа представим выражение (7.12) в виде

$$\mathscr{K}_0(i\omega)=\frac{1}{1+c(\omega)}\,e^{-i\omega t_0}, \qquad (7.13)$$

где $c(\omega)=\Omega_2(\omega)/\Omega_1(\omega)$. При $t_0=0$ оптимальный фильтр физически нереализуем, так как из-за симметрии его импульсной характеристики $g(t)$ относительно вертикальной оси, проходящей через t_0, значение $g(t)\neq 0$ при $t\leqslant 0$. Если выбрать $t_0\geqslant T$, где T — время задержки, то можно добиться, чтобы $g(t)=0$ при $t\leqslant 0$. Такой физически реализуемый фильтр будет близок к оптимальному.

Рассмотрим предельные характеристики оптимального фильтра в зависимости от отношения $c(\omega)$.

1. $c(\omega)\to 0$. В пределе это значит, что спектры сигнала и помехи не перекрываются и отношение сигнал/шум велико. В этом случае

$$\lim_{c(\omega)\to 0}\mathscr{K}_0(\omega)=1, \quad \lim_{c(\omega)\to 0}\varepsilon^2_0=\frac{1}{2\pi}\int\limits_{-\infty}^{\infty}\Omega_2(\omega)\,d\omega=\mathscr{P}_2. \qquad (7.14)$$

Оптимальный линейный фильтр является идеальным полосовым, его амплитудно-частотная характеристика (АЧХ)

$$\mathscr{K}_0(\omega) = \begin{cases} 1, & \Omega_1(\omega) \neq 0, \\ 0, & \Omega_1(\omega) = 0. \end{cases} \tag{7.15}$$

Отношение сигнал/шум на выходе $h_2 = h_{2\text{макс}} = \mathscr{P}_1/\mathscr{P}_2$.

2. $c(\omega) \to \infty$. В пределе это значит, что спектры сигнала и помехи полностью перекрываются, отношение сигнал/шум мало на всех частотах. В этом случае

$$\lim_{c(\omega)\to\infty} \mathscr{K}_0(\omega) = 0, \quad \lim_{c(\omega)\to\infty} \varepsilon^2_0 = \frac{1}{2\pi} \int_{-\infty}^{\infty} \Omega_1(\omega)\, d\omega = \mathscr{P}_1.$$

Удовлетворительное восстановление сигнала с помощью линейной фильтрации невозможно, так как квадрат минимальной среднеквадратической погрешности фильтрации в пределе равен мощности полезного сигнала и $h_2 = h_{2\text{мин}} = 1$.

3. $0 \leqslant c(\omega) \leqslant \infty$. Это наиболее общий (промежуточный) случай, когда спектры сигнала и помехи в той или иной мере перекрываются, а отношение сигнал/шум варьирует в широких пределах. В этом случае

$$0 \leqslant \mathscr{K}_0(\omega) \leqslant 1, \ \mathscr{P}_2 \leqslant \varepsilon^2_0 \leqslant \mathscr{P}_1, \quad 1 \leqslant h_2 \leqslant \mathscr{P}_1/\mathscr{P}_2. \tag{7.16}$$

Оптимальный фильтр на частотах, где $c(\omega) > 1$, ослабляет влияние помехи потому, что модуль коэффициента передачи оказывается меньшим 0,5 и с ростом $c(\omega)$ еще более уменьшается. Он обеспечивает возможно большее подавление составляющих спектра помехи и в то же время возможно меньше ослабляет (искажает) составляющие спектра сигнала.

Анализ результатов линейной фильтрации показывает, что эффективность фильтрации тем больше, чем больше отношение ширины спектра сигнала к ширине спектра помехи и чем больше отношение сигнал/шум на входе фильтра. Следовательно, для повышения эффективности линейной фильтрации целесообразно применять широкополосные сигналы и большее отношение сигнал/шум.

Зная форму спектральной плотности помехи, можно применить в передатчике предыскажение сигнала — перераспределение мощности сигнала по частотам с помощью линейного фильтра. В результате этого преобразования сигнала его мощность распределяется по полосе частот обратно пропорционально спектральной плотности помехи и эффективность оптимальной линейной фильтрации повышается. При воспроизведении сигнала в приемнике для отсутствия частотных искажений необходимо, конечно, применять предварительный фильтр с характеристикой, обратной характеристике предыскажающего фильтра.

Импульсная характеристика оптимального фильтра находится из решения интегрального уравнения

$$\int_0^t [K_1(\tau-y)+K_2(\tau-y)]\,g(y)\,dy=K_1(\tau), \tag{7.17}$$

которое называют уравнением Винера—Хопфа. Минимальная среднеквадратическая погрешность фильтрации

$$\varepsilon^2_0=\sigma^2+\int_0^t g(\tau_2)\,d\tau_2\left\{\int_0^T [K_1(\tau_2-\tau_1)+K_2(\tau_2-\tau_1)]\,g(\tau_1)\,d\tau_1-2K_1(\tau_2)\right\}, \tag{7.18}$$

где σ^2 — дисперсия $s[u(t),\,t]$.

Для иллюстрации конкретных особенностей решения (7.17) рассмотрим простейший пример линейной фильтрации гауссовского стационарного процесса с корреляционной функцией $K_1(\tau)=$ $=\sigma^2\,e^{-\alpha|\tau|}$ (см. п. 2.4.1) при условии, что $\xi(t)$ — гауссовский белый шум со спектральной плотностью $\Omega_0/2$. Уравнение (7.17) принимает вид

$$\int_0^t\left[\sigma^2 e^{-\alpha|\tau-y|}+\frac{\Omega_0}{2}\delta(\tau-y)\right]g(y)\,dy=\sigma^2\,e^{-\alpha|\tau|}. \tag{7.19}$$

Преобразуем его к виду, удобному для решения:

$$e^{-\alpha\tau}\int_0^t e^{\alpha y}g(y)\,dy+e^{\alpha\tau}\int_\tau^t e^{-\alpha y}g(y)\,dy=e^{-\alpha\tau}-\frac{\Omega_0}{2\sigma^2}g(\tau).$$

Умножив обе части этого уравнения на $e^{\alpha\tau}$ и затем дважды продифференцировав по τ, вместо интегрального уравнения получим дифференциальное уравнение второго порядка с постоянными коэффициентами:

$$g''(\tau)-\gamma^2 g(\tau)=0, \tag{7.20}$$

где $\gamma^2=\alpha^2(1+4\sigma^2/\alpha\Omega_0)$. Общее решение (7.20) имеет вид

$$g(\tau)=c_1 e^{-\gamma\tau}+c_2 e^{\gamma\tau}, \tag{7.21}$$

где коэффициенты $c_1(t)$ и $c_2(t)$ находят из системы уравнений, которая получается путем подстановки решения (7.21) в уравнение (7.19) и приравнивания коэффициентов при $e^{-\alpha\tau}$ и $e^{-\alpha(t-\tau)}$:

$$c_1(t)=\frac{(\alpha+\gamma)^2(\alpha-\gamma)}{(\alpha+\gamma)^2-(\alpha-\gamma)^2 e^{-2\gamma t}},\quad c_2(t)=\frac{(\alpha+\gamma)(\alpha-\gamma)^2}{(\alpha+\gamma)^2 e^{2\gamma t}-(\alpha-\gamma)^2}.$$

Для определения минимальной среднеквадратической погрешности оптимальную характеристику $g_0(\tau)$, найденную из (7.21),

подставляют в (7.18). После выполнения громоздких вычислений для установившегося режима получают

$$\lim_{t\to\infty} \varepsilon^2_0(t) = \Omega_0(\gamma - \alpha)/2. \tag{7.22}$$

В рассмотренном примере решение уравнения Винера—Хопфа выполняется относительно просто. Во многих случаях это решение наталкивается на серьезные трудности. Для аппаратурной реализации оптимальной характеристики фильтра необходимо воспроизводить зависимости $e^{\gamma t}$ и $e^{2\gamma t}$, что также вызывает трудности.

В 1961 г. интересные идеи были высказаны в работе Р. Е. Кальмана и Р. С. Бьюси [13]. Сущность одной из них заключается в следующем. Вместо определения частотных или импульсных характеристик линейных фильтров предложено оценивать непосредственно восстанавливаемый сигнал как решение дифференциального уравнения, коэффициенты которого определяются статистикой входных сигналов и оцениваемого. Очевидным преимуществом способа Кальмана — Бьюси является возможность относительно простого построения аналогового или цифрового вычислителя искомой оценки в зависимости от вида дифференциального уравнения. Оптимальные линейные фильтры выполняют как автоматические вычислители, с контурами обратной связи, которые включают интеграторы, цепи с переменными во времени коэффициентами передачи, сумматоры, нелинейные безынерционные устройства, объединенные таким образом, чтобы воспроизвести требуемое соотношение между входными и выходными переменными. Такие фильтры получили название фильтров Кальмана—Бьюси. Другие преимущества этих фильтров: представление процессов во временной области, что позволяет рассматривать конечные интервалы времени и выполнять фильтрацию нестационарных процессов, какими являются модулированные сигналы; ориентация на использование ЭВМ; возможность применения их в более общих задачах нелинейной фильтрации. Существенным недостатком этого способа являются трудности получения в замкнутой форме оценок, аналогичных (7.11) и (7.22), для среднеквадратических погрешностей фильтрации, иначе говоря, трудности оценки эффективности фильтрации.

7.2.2. Нелинейная фильтрация непрерывных сигналов. Основными недостатками линейной теории фильтрации являются трудности решения интегральных и дифференциальных уравнений, определяющих импульсные характеристики фильтров и восстанавливаемых сигналов, требования линейности преобразований и нормальности распределений сигналов и помех и трудности аппаратурной реализации фильтров для реальных видов модуляции, когда необходимо выделять нестационарные сигналы. Этих недостатков лишена теория нелинейной фильтрации, которая построена в основном на теории условных марковских процессов, использующей нелинейные дифференциальные уравнения [18].

Точное аналитическое решение дифференциальных уравнений нелинейной фильтрации может быть получено крайне редко, их основная ценность в том, что они позволяют непосредственно синтезировать оптимальные структурные схемы приемников, а когда распределения сигналов и помех близки к нормальным, то и оценить эффективность фильтрации.

При нелинейной фильтрации приемники являются автоматическими нелинейными устройствами, следящими за информационными параметрами сигналов, по существу, *аналоговыми или цифровыми вычислителями решения нелинейного дифференциального уравнения фильтрации.* Задачи нелинейной фильтрации так же, как и задачи линейной фильтрации нестационарных процессов, являются предметом современных исследований.

7.2.3. Цифровая фильтрация — это такая разновидность нелинейной фильтрации, в которой роль вычислителей решения нелинейного дифференциального уравнения фильтрации выполняют ЦВМ. Методы цифровой фильтрации образуют теорию цифровой обработки непрерывных сигналов. Цифровые фильтры обладают рядом преимуществ. Главное из них — возможность получения таких частотных характеристик, реализация которых с помощью обычных активных или пассивных фильтров очень сложна или просто невозможна. Другим достоинством является то, что частотные характеристики цифрового фильтра определяет всего лишь один параметр — шаг дискретизации непрерывного сигнала. Изменяя этот шаг, можно в широких пределах перестраивать фильтр. Применяя кварцевые генераторы тактовой частоты, можно обеспечить такую высокую стабильность характеристик цифрового фильтра, которая недостижима для аналоговых фильтров.

Основная трудность широкого практического применения цифровых фильтров заключается в необходимости создания быстродействующих аналого-цифровых и цифроаналоговых преобразователей, а также компактных недорогих ЦВМ. Развитие микроэлектроники и вычислительной техники показывает, что эти трудности будут преодолены.

Контрольные вопросы

1. В чем особенности определения помехоустойчивости передачи непрерывных сигналов?
2. Как ставится задача оптимального приема непрерывных сигналов?
3. Как работает квазиоптимальный следящий приемник?
4. Как определяют частотные характеристики оптимального линейного фильтра?
5. Как определяют импульсную характеристику оптимального линейного фильтра?
6. Как строят фильтры Кальмана — Бьюси?
7. В чем сущность нелинейной и цифровой фильтрации непрерывных сигналов?

7.3. ОПТИМАЛЬНЫЙ ПРИЕМ СИГНАЛОВ ПО КРИТЕРИЮ МАКСИМАЛЬНОГО ПРАВДОПОДОБИЯ

Рассмотрим особенности решения задач оптимального приема непрерывных сигналов *методом максимального правдоподобия*. Вначале решим задачу оптимальной оценки одного неизвестного параметра сигнала, а затем n параметров ортогонального разложения полезного сигнала. Конечным итогом решения задач будем считать определение отношения сигнал/шум на выходе приемника, выраженного через характеристики полезного сигнала и помехи, и оценку потенциальной помехоустойчивости передачи непрерывных сигналов.

7.3.1. Оптимальная оценка амплитуды сигнала. Такая задача возникает при измерении амплитуды сигнала или коэффициента передачи канала. Аналогично оценивают среднюю частоту или время прихода сигнала, тактовый интервал в синхронных дискретных системах связи и другие параметры. Для определенности примем, что передаваемый сигнал имел единичную амплитуду, а после прохождения по каналу его амплитуда приняла значение k, тогда $z(t)=ks(t)+\xi(t)$. Наша цель — применить метод максимального правдоподобия для оптимальной точечной оценки k.

Понятно, что точно определить k невозможно. Это обусловлено рядом причин: наличием помехи $\xi(t)$, ограниченностью времени наблюдения, наличием погрешностей измерительных приборов и т. п. Поэтому и ставится задача получить точечную оценку k^*, которая является функционалом принятой реализации $z(t)$ и в той или иной мере отражает истинное значение k. Конечно, желательно, чтобы в соответствии с основным принципом теории статистического оценивания полученная оценка была бы состоятельной, несмещенной и эффективной [9, 13, 15].

При гауссовском белом шуме $\xi(t)$ со спектральной плотностью Ω_0 функция правдоподобия имеет вид

$$f(\mathbf{z}|k)=c\exp\left\{-\frac{1}{\Omega_0}\int_0^T [z(t)-ks(t)]^2\,dt\right\}.$$

Логарифм этой функции

$$L=\ln c-\frac{1}{\Omega_0}\int_0^T [z(t)-ks(t)]^2\,dt.$$

Дифференцирование L по k и приравнивание результата нулю дает следующее *уравнение максимального правдоподобия*

$$-\frac{2k}{\Omega_0}\int_0^T [z(t)-ks(t)]\,s(t)\,dt=0. \tag{7.23}$$

Преобразуем его к виду, удобному для решения:

$$-\frac{2k}{\Omega_0}\int_0^T z(t)\,s(t)\,dt+\frac{2k^2}{\Omega_0}\int_0^T s^2(t)\,dt=0. \tag{7.24}$$

Решение уравнения

$$k^*=\frac{1}{E}\int_0^T z(t)\,s(t)\,dt. \tag{7.25}$$

Следовательно, для получения оптимальной оценки k^* необходимо применить один из видов аппаратурной реализации оптимальной линейной обработки (см. § 6.1). Например, можно использовать корреляционный приемник, в котором для упрощения схемы коэффициент усиления интегратора следует выбрать равным $1/E$. Определим качество полученной точечной оценки k^*. Подставим в (7.25) значение $z(t)$, тогда

$$k^*=\frac{1}{E}\int_0^T [ks(t)+\xi(t)]\,s(t)\,dt=k+\frac{1}{E}\int_0^T \xi(t)\,s(t)\,dt. \tag{7.26}$$

Абсолютная ошибка в оценке k

$$\varepsilon=k^*-k=\frac{1}{E}\int_0^T \xi(t)\,s(t)\,dt. \tag{7.27}$$

Найдем математическое ожидание и дисперсию ошибки

$$M[\varepsilon]=\frac{1}{E}\int_0^T M[\xi(t)]\,s(t)\,dt=0, \tag{7.28}$$

$$\mathscr{D}[\varepsilon]=M[\varepsilon^2]=\frac{1}{E^2}\,\frac{E\Omega_0}{2}=\frac{\Omega_0}{2E}. \tag{7.29}$$

Следовательно, оценка k^* является несмещенной. Так как при $T\to\infty$ и $E\to\infty$, то $\mathscr{D}[\varepsilon]\to 0$, поэтому оценка асимптотически эффективна по T. Итак, k^* можно использовать как оптимальную точечную оценку коэффициента передачи k.

7.3.2. Оптимальная оценка восстанавливаемого сигнала. Обобщим рассмотренный алгоритм получения точечной оптимальной оценки одного параметра на случай оценки n параметров ортогонального разложения полезного сигнала. По-прежнему будем считать, что принятый сигнал описывается соотношением (7.1). На интервале существования $[0, T]$ стационарный полезный сигнал, спектр которого ограничен верхней частотой F, представим

в виде канонического разложения (2.45). Положив, что постоянная составляющая $u(t)$ отсутствует, получим

$$u(t)=\sum_{k=1}^{n}\left\{\lambda_{2k}\sqrt{2}\cos\left\{\frac{2\pi}{T}kt\right\}+\right.$$

$$\left.+\lambda_{2k-1}\sqrt{2}\sin\left\{\frac{2\pi}{T}(2k-1)t\right\}\right\}=\sum_{k=1}^{n}\lambda_k\varphi_k(t), \qquad (7.30)$$

где λ_k — случайные параметры $u(t)$, точечные оценки которых необходимо получить в процессе решения оптимальной задачи обработки непрерывного сигнала; $\varphi_k(t)$ — ортонормированные гармонические функции; $n=2FT$ — база $u(t)$.

Введем для упрощения записи вектор-параметр $\lambda=(\overline{\lambda_1, \lambda_n})$, тогда принимаемый сигнал можно представить в виде

$$z(t)=s(\lambda, t)+\xi(t). \qquad (7.31)$$

Задача оценки $u(t)$ сводится теперь к задаче точечной оценки вектор-параметра λ или, иначе говоря, совокупности случайных величин $\lambda_1, \lambda_2, \ldots, \lambda_n$. Качество оценки $u(t)$ определяется совместной оценкой n параметров — координат вектора λ. *Совместные максимально правдоподобные оценки* определяются из решения системы уравнений правдоподобия

$$\partial[\ln f(\mathbf{z}\,|\,\lambda)]/\partial\lambda_k=0, \quad k=\overline{1, n}. \qquad (7.32)$$

При флуктуационной помехе в гауссовом канале и известной аналитической форме сигнала-переносчика $s(t)$ функция правдоподобия

$$f(\mathbf{z}\,|\,\lambda)=c\exp\left\{-\frac{1}{\Omega_0}\int_0^T[z(t)-s(\lambda, t)]^2\,dt\right\}. \qquad (7.33)$$

Анализ (7.33) показывает, что максимально правдоподобные оценки при флуктуационном шуме в канале минимизируют функционал ε^2, т. е. обеспечивают

$$\min_{\lambda}\varepsilon^2=\min_{\lambda}\frac{1}{\Omega_0}\int_0^T[z(t)-s(\lambda, t)]^2\,dt.$$

Следовательно, при оптимальном приеме непрерывных сигналов, как и в случае приема дискретных сигналов, приемник выделяет в результате вычисления оптимальных λ^* тот полезный сигнал $s^*(\lambda^*, t)$, который в гильбертовом пространстве ближе всего к принятому $z(t)$.

Если помеха отсутствует, то $\min\limits_{\lambda}\varepsilon^2=0$, и такой приемник точно восстанавливает полезные сигналы на выходе детектора.

Последовательное логарифмирование (7.33), дифференцирование по λ_k и приравнивание результата нулю дает систему уравнений максимального правдоподобия:

$$-\frac{2}{\Omega_0}\int_0^T [z(t)-s(\overline{\lambda_1,\ \lambda_n};\ t)]\frac{\partial s(\lambda_1,\ \lambda_n;\ t)}{\partial\lambda_k}dt=0. \qquad (7.34)$$

Решив эту систему, получим оценки $\overline{\lambda^*_1,\ \lambda^*_n}$, оптимальные по критерию максимального правдоподобия. Анализ показывает, что при определенных свойствах случайных процессов $z(t)$, на практике обычно имеющих место, это решение является единственным и дает совместно состоятельные, несмещенные и асимптотически эффективные по T точечные оценки $\overline{\lambda^*_1,\ \lambda^*_n}$ [13, 15].

Максимально правдоподобная оценка восстанавливаемого сигнала

$$u^*(t)=\sum_{k=1}^{n}\lambda^*_k\varphi_k(t)=\sum_{k=1}^{n}(\lambda_k+\Delta\lambda_k)\varphi_k(t)=u(t)+\sum_{k=1}^{n}\Delta\lambda_k\varphi_k(t), \qquad (7.35)$$

где $\Delta\lambda_k$ — асимптотически нормально распределенные случайные величины, как будет показано далее, с нулевыми средними значениями. Погрешность оценки полезного сигнала

$$\varepsilon(t)=u^*(t)-u(t)=\sum_{k=1}^{n}\Delta\lambda_k\varphi_k(t) \qquad (7.36)$$

можно рассматривать как шум на выходе приемника, а дисперсию $\mathcal{D}[\sqrt{2}\Delta\lambda_k]=2\mathcal{D}$ — как мощность этого шума на частоте f_k. Для определения дисперсии $\mathcal{D}[\Delta\lambda_k]$ выразим $\Delta\lambda_k$ через характеристики передаваемого сигнала и помехи. Допустим, что под действием слабой помехи $\xi(t)$ колебание $z(t)$ получит малое приращение, равное этой помехе, т. е.

$$\Delta z(t)=\xi(t). \qquad (7.37)$$

Тогда координаты $\overline{\lambda_1,\ \lambda_k}$ полезного сигнала на выходе приемника получат приращения $\Delta\lambda_j$, которым соответствует приращение канального сигнала $\Delta s(\Delta\lambda,\ t)$, определяемое соотношением

$$\Delta s(\Delta\lambda,\ t)=\sum_{j=1}^{n}\frac{\partial s(\lambda,\ t)}{\partial\lambda_j}\Delta\lambda_j. \qquad (7.38)$$

Средний квадрат отклонения между колебанием $z(t)+\Delta z(t)$ и сигналом $s(\lambda,\ t)+\Delta s(\Delta\lambda,\ t)$

$$\varepsilon^2=\int_0^T[z(t)+\Delta z(t)-s(\lambda,\ t)-\Delta s(\Delta\lambda,\ t)]^2dt. \qquad (7.39)$$

Дифференцируя (7.39) по λ_k, получаем следующие уравнения максимального правдоподобия

$$\int_0^T [z(t) - s(\lambda, t)] \frac{\partial s(\lambda, t)}{\partial \lambda_k} dt = \int_0^T [\Delta z(t) - \Delta s(\Delta\lambda, t)] \frac{\partial s(\lambda, t)}{\partial \lambda_k} dt, \tag{7.40}$$

где производная $\partial \Delta s(\Delta\lambda, t)/\partial\lambda_k$ и порождаемые ею составляющие не учитываются как величины второго порядка малости.

Учитывая (7.37), (7.38), (7.40) можем записать в виде

$$\frac{1}{T}\int_0^T \xi(t) \frac{\partial s(\lambda, t)}{\partial \lambda_k} dt = \sum_{j=1}^{n} \Delta\lambda_j \frac{1}{T}\int_0^T \frac{\partial s(\lambda, t)}{\partial \lambda_j}\frac{\partial s(\lambda, t)}{\partial \lambda_k} dt. \tag{7.41}$$

Используя основное свойство ортогональных разложений, можно показать, что для всех реально применяемых видов модуляции функции $\partial s(\lambda, t)/\partial\lambda_j$ и $\partial s(\lambda, t)/\partial\lambda_k$ взаимно ортогональны на интервале $[0, T]$. Поэтому

$$\Delta\lambda_k = \frac{1}{\langle[\partial s(\lambda, t)/\partial\lambda_k]^2\rangle}\frac{1}{T}\int \xi(t)\frac{\partial s(\lambda, t)}{\partial \lambda_k} dt.$$

Выполнив операции усреднения по множеству и по времени, получим

$$M[\Delta\lambda_k] = 0, \ \mathscr{D}[\Delta\lambda_k] = M[\Delta\lambda^2{}_k] = \frac{\Omega_0}{2T\left\langle\left[\dfrac{\partial s(\lambda, t)}{\partial\lambda_k}\right]^2\right\rangle}. \tag{7.42}$$

Величину $\mathscr{D}[\Delta\lambda_k] = \sigma^2{}_{2k-1} = \sigma^2{}_{2k} = \sigma^2{}_k$ в дальнейшем будем использовать как оценку дисперсии параметров $\lambda^*{}_k$ канонического разложения $u(t)$ (7.30). Поэтому мощность шума на выходе детектора на частоте f_k

$$\sigma^2(f_k) = M[\sqrt{2}\lambda^*{}_{2k}]^2 = M[\sqrt{2}\lambda^*{}_{2k-1}]^2 = 2\sigma^2{}_k.$$

Следовательно,

$$\sigma^2(f_k) = \frac{\Omega_0}{T\left\langle\left[\dfrac{\partial s(\lambda, t)}{\partial\lambda_k}\right]^2\right\rangle} = \frac{\Omega_0\Delta f}{\left\langle\left[\dfrac{\partial s(\lambda, t)}{\partial\lambda_k}\right]^2\right\rangle}, \tag{7.43}$$

так как смежные спектральные составляющие разложения (7.30) сдвинуты по частоте на интервал $\Delta f = 1/T$.

Из (7.43) следует, что спектральная плотность шума на выходе детектора

$$\Omega(f) = \frac{\Omega_0}{\langle[\partial s(\lambda, t)/\partial\lambda_k]^2\rangle}. \tag{7.44}$$

Отношение средней мощности сигнала к средней мощности шума на выходе оптимального приемника

$$h_2=\left(\frac{\mathcal{P}_1}{\mathcal{P}_2}\right)_2=\mathcal{P}_{12}\Bigg/\int\limits_0^{\Delta F}\Omega(f)\,df, \tag{7.45}$$

где $\mathcal{P}_{12}$ — средняя мощность $u(t)$.

7.3.3. Оценка потенциальной помехоустойчивости передачи непрерывных сигналов. Для оценки потенциальной помехоустойчивости непрерывных приемников широкое распространение получила такая характеристика, как выигрыш (см. § 5.6)

$$\mathfrak{y}=\frac{(\mathcal{P}_1/\mathcal{P}_2)_2}{(\mathcal{P}_1/\mathcal{P}_2)_1}=\frac{h_2}{h_1}, \tag{7.46}$$

где $(\mathcal{P}_1/\mathcal{P}_2)_1$ — отношение сигнал/шум на входе приемника (в канале).

По аналогии с (6.10) ее можно рассматривать как эффективность оптимальной обработки непрерывных сигналов или в более общем применении как эффективность непрерывных систем связи. Это обусловлено тем, что, как бы ни выбиралась количественная мера верности передачи информации, она является возрастающей функцией этого отношения. Другое дело, что верность определяется не только одним этим отношением, а и рядом других факторов: шириной спектров сигналов и помех, характером распределений амплитуд и фаз сигналов и помех, способом регистрации сигналов и т. п.

Чтобы учесть различную ширину спектров сигналов и помех, отношение (7.46) нормируют с помощью коэффициента частотной избыточности модуляции $\nu_1=\Delta F_1/\Delta F$, равного отношению полосы пропускания входных цепей приемника (ширины спектра модулированного сигнала) к полосе частот полезного сигнала. Тогда обобщенный выигрыш (обобщенная оценка эффективности)

$$\mathfrak{y}_1=\frac{\mathfrak{y}}{\nu_1}=\frac{(\mathcal{P}_1/\mathcal{P}_2)_2\,\Delta F}{(\mathcal{P}_1/\mathcal{P}_2)_1\,\Delta F_1}=\frac{h_{21}}{h_{11}}. \tag{7.47}$$

Для удобства сравнения различных модулирующих сигналов их нормируют, чтобы

$$u_{\text{макс}}(t)=\max|u(t)/u_m|=1, \tag{7.48}$$

где u_m — максимальное значение $u(t)$, и вводят пик-фактор полезного сигнала

$$\Pi=\frac{u_{\text{макс}}(t)}{\sqrt{\frac{1}{T}\int\limits_0^T u^2(t)\,dt}}=\frac{u_{\text{макс}}(t)}{\sqrt{\mathcal{P}_{12}}}=\frac{1}{\sqrt{\mathcal{P}_{12}}}. \tag{7.49}$$

Из (7.49) следует, что

$$\mathcal{P}_{12}=1/\Pi^2, \tag{7.50}$$

а отношение (7.45) принимает вид

$$\left(\frac{\mathcal{P}_1}{\mathcal{P}_2}\right)_2 = \frac{1}{\Pi^2 \int\limits_0^{\Delta E} \Omega(f)\, df} = \left[\Pi^2 \int\limits_0^{\Delta F} \frac{\Omega_0 df}{\langle [\partial s(\lambda,\ t)/\partial\lambda_k]^2 \rangle}\right]^{-1}. \tag{7.51}$$

Таким образом, применение метода максимального правдоподобия позволяет получить оценки максимального отношения сигнал/шум на выходе оптимального приемника и по максимальному выигрышу $\mathfrak{y}$ (7.46) или по максимальному обобщенному выигрышу $\mathfrak{y}_1$ (7.47) оценить потенциальную помехоустойчивость различных методов модуляции.

Контрольные вопросы

1. Для чего применяют ортогональное разложение полезных сигналов при рассмотрении задач обработки непрерывных сигналов?
2. Как методом максимального правдоподобия получают точечную оценку одного параметра полезного сигнала?
3. Как используют метод максимального правдоподобия для оптимального восстановления сигнала из аддитивного шума?
4. Как определяют дисперсии коэффициентов ортогонального разложения полезного сигнала?
5. Каким соотношением определяется мощность шумов и спектральная плотность сигнала на выходе оптимального приемника?
6. Как определяют выигрыш и обобщенный выигрыш непрерывных систем связи?
7. Как отношение сигнал/шум на выходе приемника связано с пик-фактором полезного сигнала, полосой спектра полезного сигнала и спектральной плотностью шумов?
8. Почему отношение сигнал/шум на выходе приемника можно использовать для оценки потенциальной помехоустойчивости непрерывных систем связи?

7.4. ПОТЕНЦИАЛЬНАЯ ПОМЕХОУСТОЙЧИВОСТЬ МЕТОДОВ МОДУЛЯЦИИ

Основным соотношением для определения потенциальной помехоустойчивости методов модуляции является выражение (7.51), в котором величина $\langle [\partial s(\lambda,\ t)/\partial\lambda_k]^2 \rangle$ полностью определяется характером оператора модуляции M. Оператор ψ в приближенных оценках обычно не учитывают. В соответствии с классификацией Котельникова различают *прямые и непрямые методы модуляции*. Прямыми методами аналоговой модуляции являются амплитудная (АМ), однополосная (ОМ) и фазовая модуляция (ФМ). В этих методах модуляции полезный сигнал без всяких преобразований входит в аналитическое выражение $s[u(t),\ t]$. К непрямому методу относится частотная модуляция (ЧМ), при которой полезный сигнал входит под знаком интеграла в выражение для $s[u(t),\ t]$, поэтому ее часто называют интегральным методом модуляции.

Помимо этого, целесообразно различать *линейные и нелинейные методы модуляции* — по характеру преобразования $u(t)$ в $s[u(t),\ t]$ с помощью оператора M. По этому признаку однопо-

лосная модуляция является линейным видом, а частотная и фазовая — нелинейными. В последних модулированный сигнал $s[u(t), t]$ получают путем нелинейного (тригонометрического) преобразования полезного сигнала $u(t)$.

Задача определения потенциальной помехоустойчивости различных методов модуляции сводится к определению значения $\langle[\partial s(\lambda, t)/\partial\lambda_k]^2\rangle$, вычислению отношения (7.51), оценке выигрыша (7.46) и обобщенного выигрыша (7.47). Определим потенциальную помехоустойчивость двух методов модуляции: АМ и ЧМ. Изучим особенности определения потенциальной помехоустойчивости для прямого (АМ) и непрямого (ЧМ) методов модуляции, для линейного (ОМ) и нелинейного (ЧМ). Исследование потенциальной помехоустойчивости других методов модуляции выполняют аналогично, по такому же алгоритму.

7.4.1. Потенциальная помехоустойчивость прямых методов модуляции. Для прямых методов модуляции

$$\left\langle\left[\frac{\partial s(\lambda, t)}{\partial\lambda_k}\right]^2\right\rangle=\left\langle\left\{\frac{\partial s[u(t), t]}{\partial u}\frac{\partial u}{\partial\lambda_k}\right\}^2\right\rangle=\left\langle\left\{\frac{\partial s[u(t), t]}{\partial u}\varphi_k(t)\right\}^2\right\rangle. \quad (7.52)$$

Так как спектр функции $(\partial s/\partial u)^2$ обычно лежит в области значительно более высоких частот, чем спектр полезного сигнала $u(t)$ (выше частоты гармонического колебания $\varphi_k(t)$), то в силу ортогональности функций с неперекрывающимися спектрами среднее значение их произведения равно произведению средних значений. Следовательно,

$$\left\langle\left\{\frac{\partial s[u(t), t]}{\partial u}\right\}^2\right\rangle\langle\varphi^2_k(t)\rangle=\left\langle\left\{\frac{\partial s[u(t), t]}{\partial u}\right\}^2\right\rangle, \quad (7.53)$$

где в силу ортонормированности функций $\varphi_k(t)$ значение $\langle\varphi^2_k(t)\rangle = 1$.

Функция (7.53) от частоты не зависит, поэтому спектральная плотность шума на выходе оптимального приемника равномерна и определяется соотношением

$$\Omega_2=\frac{\Omega_0}{\langle\{\partial s[u(t), t]/du\}^2\rangle}. \quad (7.54)$$

Эффективность прямых методов модуляции найдем, используя (7.51), (7.46), (7.47):

$$\mathfrak{y}=\frac{\Delta F_1\langle\{\partial s[u(t), t]/du\}^2\rangle}{\Delta F\Pi^2\mathcal{P}_1}, \quad \mathfrak{y}_1=\frac{\langle\{\partial s[u(t), t]/\partial u\}^2\rangle}{\Pi^2\mathcal{P}_1}, \quad (7.55)$$

где $\mathcal{P}_1$ — мощность модулированного сигнала.

Рассмотрим оценку *потенциальной помехоустойчивости АМ*. В этом случае $s[u(t), t]=a[1+mu(t)]\cos(\omega t+\varphi)$. Производная

$$\partial s[u(t), t]/\partial u=am\cos(\omega t+\varphi),$$

среднее значение квадрата этой производной

$$\langle a^2 m^2 \cos^2(\omega t+\varphi)\rangle = a^2 m^2/2.$$

Средняя мощность модулированного сигнала

$$\mathscr{P}_1 = \langle \{s[u(t),\ t]\}^2\rangle = \frac{a^2}{2}[1+m^2\langle u^2(t)\rangle] = \frac{a^2}{2}\left(1+\frac{m^2}{\Pi^2}\right).$$

Так как при АМ $\Delta F_1 = 2\Delta F$, то

$$\mathfrak{y}_{\text{АМ}} = \frac{\Delta F_1}{\Delta F}\,\frac{m^2}{m^2+\Pi^2} = \frac{2m^2}{m^2+\Pi^2};\quad \mathfrak{y}_{1\text{АМ}} = \frac{m^2}{m^2+\Pi^2}. \tag{7.56}$$

Из (7.56) следует, что $\mathfrak{y}_{\text{АМ}}$ и $\mathfrak{y}_{1\text{АМ}}$ определяются только коэффициентом модуляции и пик-фактором. Значения коэффициентов эффективности меньше единицы, так как $\Pi>1$, $m<1$. Например, для речевого сигнала $\Pi \simeq 3$, $m=0{,}8$ и $\mathfrak{y}_{\text{АМ}}=0{,}068$, $\mathfrak{y}_{1\text{АМ}}=0{,}034$.

Следовательно, при АМ необходимо говорить не о выигрыше в отношении сигнал/шум, а о проигрыше. Проигрыш обусловлен тем, что для передачи используется только часть мощности модулированного сигнала. Аналогичные расчеты для ОМ и амплитудно-импульсной модуляции (АИМ) дают

$$\mathfrak{y}_{\text{ОМ}} = \mathfrak{y}_{1\text{ОМ}} = 1,\quad \mathfrak{y}_{\text{АИМ}} = \nu_1 \approx 1/\tau F,\quad \mathfrak{y}_{1\text{АИМ}} = 1,$$

где τ — длительность импульсов.

Интересно отметить, что из-за линейной зависимости между модулированным сигналом $s[u(t),\ t]$ и полезным сигналом $u(t)$ для АМ и ОМ полученные соотношения справедливы как при слабых, так и при сильных помехах. Для нелинейных методов модуляции этот вывод несправедлив.

7.4.2. Потенциальная помехоустойчивость непрямых методов модуляции. Для определенности рассмотрим интегральную модуляцию. В этом случае $s[\nu(t),\ t] = s\{[\int u(t)\,dt],\ t\}$. Следовательно,

$$\nu(t) = \sum_{k=1}^{n} \lambda_k \int \varphi_k(t)\,dt. \tag{7.57}$$

Среднее значение

$$\langle[\partial\nu(t)/\partial\lambda_k]^2\rangle = \left\langle\left[\int \varphi_k(t)\,dt\right]^2\right\rangle = 1/\omega^2_k,$$

так как $\varphi_k(t)$ — гармонические функции с частотой ω_k (7.30) и

$$\int \varphi_k(t)\,dt = \begin{cases} -(\sqrt{2}/\omega_k)\cos\omega_k t, \\ (\sqrt{2}/\omega_k)\sin\omega_k t. \end{cases}$$

Математическое ожидание

$$\left\langle\left\{\frac{\partial s[\nu(t),\ t]}{\partial\lambda_k}\right\}^2\right\rangle = \left\langle\left\{\frac{\partial s[\nu(t),\ t]}{\partial\nu}\right\}^2\right\rangle\left\langle\left(\frac{\partial\nu}{\partial\lambda_k}\right)^2\right\rangle =$$

$$= \frac{1}{\omega^2_k}\left\langle\left\{\frac{\partial s[\nu(t),\ t]}{\partial\nu}\right\}^2\right\rangle,$$

где второй сомножитель от частоты не зависит.

Спектральная плотность шума на выходе оптимального приемника при интегральной модуляции

$$\Omega(f) = \frac{(2\pi f)^2 \Omega_0}{\langle \{\partial s[\nu(t), t]/\partial\nu\}^2\rangle}, \tag{7.58}$$

т. е. является параболической функцией частоты. Это важный вывод потому, что *при многоступенчатых методах модуляции*, например при двухступенчатой модуляции типа ЧМ-ОМ, второй детектор приемника работает в режиме выделения полезного сигнала из аддитивного шума со спектром (7.58). Помеха уже не является белым шумом и все полученные соотношения требуют уточнения.

Выигрыш и обобщенный выигрыш при интегральной модуляции и оптимальном приеме в соответствии с (7.46), (7.47)

$$\mathfrak{y} = \frac{\Delta F_1 \left\langle \left\{ \frac{\partial s[\nu(t), t]}{\partial\nu} \right\}^2 \right\rangle}{\Pi^2 \mathscr{P}_1 (2\pi)^2 \int\limits_0^{\Delta F} f^2 df} = \frac{3\Delta F_1 \left\langle \left\{ \frac{\partial s[\nu(t), t]}{\partial\nu} \right\}^2 \right\rangle}{\Pi^2 \mathscr{P}_1 (2\pi)^2 \Delta F^3}, \tag{7.59}$$

$$\mathfrak{y}_1 = 3 \left\langle \left\{ \frac{\partial s[\nu(t), t]}{\partial\nu} \right\}^2 \right\rangle \Big/ \Pi^2 \mathscr{P}_1 (2\pi)^2 \Delta F^2. \tag{7.60}$$

Рассмотрим оценку *потенциальной помехоустойчивости ЧМ.* В этом случае модулированный сигнал $s[\nu(t), t] = a\cos[\omega t + \Delta\omega\nu(t) + \varphi]$, где $\Delta\omega = 2\pi\Delta f$ — девиация частоты; $\nu(t) = \int u(t)\,dt$.

Тогда

$$\langle \{\partial s[\nu(t), t]/\partial\nu\}^2\rangle = a^2\Delta\omega^2/2.$$

Следовательно, *выигрыш ЧМ*

$$\mathfrak{y}_{\text{ЧМ}} = 3\Delta f^2 \Delta F_1/\Pi^2\Delta F^3 = 3\beta^2_{\text{ЧМ}}\Delta F_1/\Pi^2\Delta F, \tag{7.61}$$

$$\mathfrak{y}_{1\text{ЧМ}} = 3\beta^2_{\text{ЧМ}}/\Pi^2, \tag{7.62}$$

где $\beta_{\text{ЧМ}}$ — индекс ЧМ. Из (7.62) следует вывод о том, что высокая потенциальная помехоустойчивость ЧМ ($\mathfrak{y}_{\text{ЧМ}} \gg 1$) обеспечивается при фиксированном Π благодаря частотной избыточности, так как $\nu_1 = \Delta F_1/\Delta F \gg 1$ и $\beta^2_{\text{ЧМ}} \gg 1$. Ширина спектра ЧМ колебания $\Delta F_1 \approxeq 2\Delta F\beta_{\text{ЧМ}}$

Для сравнения укажем *выигрыш ФМ и ФИМ*

$$\mathfrak{y}_{\text{ФМ}} = \frac{\Delta F_1}{\Delta F}\,\frac{\beta^2_{\text{ФМ}}}{\Pi^2}, \quad \mathfrak{y}_{\text{ФИМ}} \approxeq \frac{0{,}6\nu^3_1}{\Pi^2}, \tag{7.63}$$

$$\mathfrak{y}_{1\text{ФМ}} = \frac{\beta^2_{\text{ФМ}}}{\Pi^2} = \frac{\Delta F^2_1}{4\Pi^2\Delta F^2} = \frac{\nu^2_1}{4\Pi^2}, \quad \mathfrak{y}_{1\text{ФИМ}} \approxeq \frac{0{,}6\nu^2_1}{\Pi^2}, \tag{7.64}$$

где $\beta_{\text{ФМ}} = \Delta\varphi$ — индекс фазовой модуляции. Так же как и при ЧМ, выигрыш ФМ обусловлен частотной избыточностью.

Следует отметить, что расширение спектра модулированных сигналов имеет и отрицательные последствия. При селективных замираниях (см. § 4.1) имеют место частотные искажения модулированных сигналов, что приводит к значительному снижению реальной помехоустойчивости по сравнению с потенциальной. Поэтому для непрерывных каналов с селективными замираниями предпочтение отдают не ЧМ и ФМ, а ОМ.

При сильных помехах (слабых сигналах) полученные для интегральной модуляции отношения несправедливы. Вызвано это тем, что, во-первых, условие (7.37) не выполняется, а во-вторых, имеет место пороговый эффект из-за нелинейности преобразования $u(t)$ в $s[u(t), t]$.

Контрольные вопросы

1. Какое соотношение является основным для определения потенциальной помехоустойчивости непрерывных систем?
2. Как классифицируют методы модуляции?
3. Каков алгоритм определения потенциальной помехоустойчивости передачи непрерывных сигналов?
4. Как определяется потенциальная помехоустойчивость прямых методов модуляции?
5. Как определяется потенциальная помехоустойчивость АМ?
6. Как определяется потенциальная помехоустойчивость непрямых методов модуляции?
7. Как определяется потенциальная помехоустойчивость ЧМ?
8. Сравните помехоустойчивость АМ и ЧМ.

7.5. ПОРОГОВЫЙ ЭФФЕКТ НЕЛИНЕЙНЫХ МЕТОДОВ МОДУЛЯЦИИ

Пороговый эффект проявляется для всех нелинейных методов модуляции в том, что, начиная с некоторого «порогового» отношения сигнал/шум на входе приемника, потенциальная помехоустойчивость модуляции резко падает и становится хуже помехоустойчивости линейных видов модуляции. Природа порогового эффекта обусловлена тем, что из-за нелинейности обратного преобразования «смеси» модулированного сигнала и шума в копию полезного сигнала при больших сигналах на входе отношение (7.46) определяется мощностью сигналов, а при больших помехах — мощностью помех. При определенном значении h^*_1, которое называют пороговым, возникает точка перегиба на графике зависимости h_2 от h_1 и h_2 уменьшается более резко по сравнению с линейными видами модуляции.

Рис. 7.3 иллюстрирует появление порогового эффекта при ФМ. На рисунке приняты следующие обозначения: $A(t)$, $\Phi(t)$ — огибающая и фаза модулированного сигнала, $\xi_1(t)$ — огибающая малой помехи, $\xi_2(t)$ — огибающая большой помехи, $\Delta\Phi_1(t)$, $\Delta\Phi_2(t)$ — приращение фазы модулированного сигнала под воздействием помех. Когда помеха мала, отношение $\Delta\Phi_1(t)/\Phi(t) \ll 1$ для всех моментов времени. С ростом отношения сигнал/шум на входе растет и отношение сигнал/шум на выходе. При большой помехе

$\Delta\Phi_2(t)/\Phi(t) \gg 1$ для всех моментов времени, полезный сигнал «подавлен» помехой и выделить его невозможно. Порог формируется при таком отношении

$$\Delta\Phi(t)/\Phi(t) = \delta, \tag{7.65}$$

при котором начинается уверенное (с большой вероятностью) выделение полезного сигнала из шума.

Следует специально оговорить, что увеличение частотной избыточности (расширение полосы спектра ЧМ и ФМ сигналов) не приводит к невозможности рассмотрения модулированного сигнала как узкополосного процесса. Это обусловлено тем, что при любых $\nu_1 = \Delta F_1/\Delta F$ отношение $\Delta F_1/F_0$, где F_0 — средняя частота модулированного процесса, для практически важных случаев всегда намного меньше единицы.

Рис. 7.3. Иллюстрация появления порогового эффекта

7.5.1. Оценка порогового значения h^*_1. Оценку «снизу» порога h^*_1 получают, используя выражение (5.89) для идеальной модуляции (идеальной системы)

$$[1 + (\mathscr{P}_1/\mathscr{P}_2)_1]^{\nu_1} - 1 = \mathfrak{y}_0 (\mathscr{P}_1/\mathscr{P}_2)_1. \tag{7.66}$$

Если частотная избыточность отсутствует, что имеет место для ОМ, то $\mathfrak{y}_0 = 1$, $\nu_1 = 1$ и уравнение (7.66) вырождается в тождество. Это значит, что пороговый эффект отсутствует и (7.66) удовлетворяется при произвольных отношениях h_1.

Если $\nu_1 > 1$ и $\mathfrak{y}_0 > 1$, решение (7.66) существует. Пороговое значение h^*_1 можно найти из решения приближенного уравнения

$$\nu_1 \log(1 + h^*_1) = \log h^*_1 + \log \mathfrak{y}_0. \tag{7.67}$$

Рассмотрим ФМ с $\nu_1 = 50$ и $\Pi = 3$ (модуляция речевым сигналом), тогда (7.67) имеет вид

$$50 \log(1 + h^*_1) = \log h^*_1 + \log(50/4 \cdot 3^2),$$

отсюда $h^*_1 = 0{,}13$.

В реальных системах пороговый эффект проявляется раньше. Рассмотрим, как приближенно можно определить h^*_1 для реальных систем. Если помеха является белым шумом, то ее огибающая, как показано в § 2.7, подчиняется распределению Релея

$$f_1(\xi) = (\xi/\sigma^2) \exp(-\xi^2/2\sigma^2),$$

где $\sigma^2 = \mathscr{P}_2$ — мощность помехи на входе приемника. *Пороговое значение амплитуды* A^* полезного сигнала найдем из условия уверенного выделения сигнала из шума. Выберем вероятность того, что помеха превысит сигнал, равный 10^{-3}, тогда

$$P(\xi < A) = \int_0^A f_1(\xi)\, d\xi = 0{,}999. \tag{7.68}$$

Вычислив интеграл (7.78), получим

$$\exp\left(-\frac{A^2}{2\sigma^2}\right)=10^{-3}, \quad h^*{}_1=\frac{A^2}{2\sigma^2}\cong 7, \quad A^*=\sqrt{\frac{7}{2}}\,\sigma. \qquad (7.69)$$

7.5.2. Связь $h^*{}_1$ и ΔF, β, Ω_0. Формулы (7.68) и (7.69) определяют примерное значение порога для ЧМ и ФМ. Определим, как *пороговое значение мощности сигнала* зависит от спектральной плотности помехи, ширины спектра сигнала и индекса модуляции. Если спектральная плотность помехи Ω_0 постоянна, мощность шума на входе приемника $P_2=\Omega_0\Delta F_1$. Величина ΔF_1 зависит от индекса модуляции и ширины спектра полезного сигнала при больших индексах модуляции следующим образом:

$$\Delta F_1 \simeq 2\Delta F\beta. \qquad (7.70)$$

Поэтому с учетом (7.69)

$$\mathscr{P}^*{}_1=7\mathscr{P}_2\simeq 14\Omega_0\Delta F\beta. \qquad (7.71)$$

Из соотношения (7.71) следует, что с ростом β растет и пороговое значение мощности сигнала. *Пороговое значение индекса модуляции* при заданном отношении сигнал/шум на входе приемника

$$\beta^*\simeq\mathscr{P}^*{}_1/14\Omega_0\Delta F\simeq\mathscr{P}^*{}_1/14\mathscr{P}^*{}_2\simeq h^*{}_1/14. \qquad (7.72)$$

Следовательно, зная характеристики канала и мощность передатчика (отношение сигнал/шум на входе приемника), индексы модуляции необходимо выбирать из условия

$$\beta<\beta^*. \qquad (7.73)$$

На рис. 7.4 показаны графики, иллюстрирующие пороговый эффект и условие (7.73). В качестве эталонной используется зависимость $h_2(h_1)$ для ОМ. Кривые для ЧМ построены при $\beta_2<\beta_3<$ $<\beta_4$. Точки излома графиков соответствуют пороговым значениям $h^*{}_{1i}$ и $\beta^*{}_i$.

7.5.3. Снижение пороговой мощности сигнала. Для снижения пороговой мощности сигнала при заданной помехе применяют ряд методов. Большинство из них основано на принципе «сжатия» спектра сигнала в приемнике путем слежения за текущей шириной спектра модулированного сигнала. Текущая ширина спектра примерно в β раз меньше ΔF_1. Это объясняется статистическими особенностями реальных полезных сигналов. Например, общая полоса ΔF спектра речевого сигнала выбирается обычно равной 3400 Гц, однако в каждый момент времени присутствуют не все составляющие спектра в пределах этой полосы, а только некоторые из них, они и участвуют в модуляции. Из-за этого реальная спектральная плотность модулированного сигнала «плавает» во времени по полосе частот в пределах общей полосы модулированного сигнала. Слежение за текущей средней частотой узкополосным приемником достигается путем введения обратной связи по частоте или применением следящего фильтра промежуточной частоты.

Для иллюстрации особенностей аппаратурной реализации методов снижения пороговой мощности передатчика или, что то же, увеличения дальности связи при фиксированной мощности передатчика, рассмотрим *структурную схему приемника ЧМ сигналов с обратной связью по частоте* (рис. 7.5) [2]. Обратная связь обеспечивает изменение частоты гетеродина Γ, синхронное с изменением частоты принимаемого сигнала. Если коэффициент передачи прямого тракта имеет величину $\mathscr{K}$, а коэффициент передачи цепи

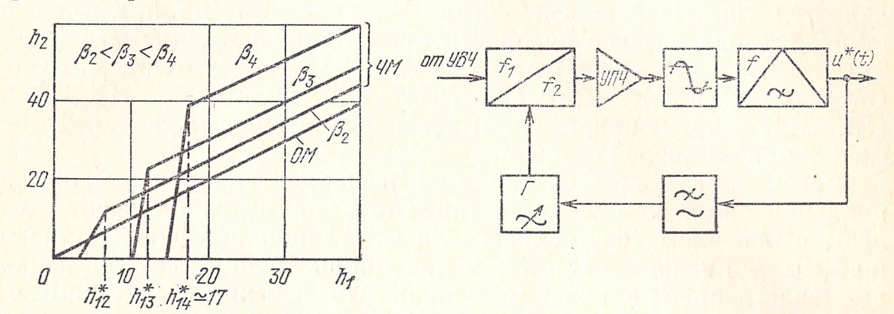


Рис. 7.4. Зависимость h_2 от h_1 при ОМ и ЧМ

Рис. 7.5. Фрагмент схемы приемника ЧМ сигналов с обратной связью по частоте

обратной связи γ_0, то параметры ЧМ сигнала на выходе усилителя промежуточной частоты ΔF_{12} и β_2 можно выразить через параметры входного сигнала ΔF_1 и β_1 следующим образом:

$$\Delta F_{12}=\Delta F_1/(1+\mathscr{K}\gamma_0),\ \ \beta_2=\beta_1/(1+\mathscr{K}\gamma_0). \qquad (7.74)$$

Следовательно и полоса, и девиация сигнала на выходе УПЧ в $(1+\mathscr{K}\gamma_0)$ раз меньше, чем на входе приемника. Поэтому полосу пропускания УПЧ можно соответственно уменьшить, мощность шумов на входе детектора упадет, что согласно (7.71) влечет за собой и уменьшение пороговой мощности сигнала. Так как

$$\Delta F_1=2\Delta F\beta_1,$$

то $\Delta F_{12}=2\Delta F\beta_1/(1+\mathscr{K}\gamma_0)$. (7.75)

При

$$\beta_1/(1+\mathscr{K}\gamma_0)\simeq 1 \quad \Delta F_{12}\simeq 2\Delta F \qquad (7.76)$$

и в пределе порог помехоустойчивости может быть снижен в β_1 раз. Эксперименты показывают, что реально добиваются снижения порога на 5—9 дБ.

Контрольные вопросы

1. Как проявляется пороговый эффект?
2. Как приближенно оценить значение порога?
3. Как связано пороговое значение сигнала со спектральной плотностью помехи, шириной спектра полезного сигнала и индексом девиации?
4. Как определить пороговое значение индекса девиации?
5. Какие методы применяют для снижения пороговой мощности сигнала?
6. Какой предельный выигрыш можно получить в приемнике со следящей обратной связью по частоте сигнала?

7.6. ПОТЕНЦИАЛЬНАЯ ПОМЕХОУСТОЙЧИВОСТЬ МНОГОСТУПЕНЧАТЫХ МЕТОДОВ МОДУЛЯЦИИ

Во многих реальных системах связи (системы радиосвязи, многоканальные системы, телеметрические и др.) канальный сигнал образуется в результате последовательного применения нескольких методов модуляции (многоступенчатая модуляция). Модулирующим сигналом для каждой следующей ступени модуляции является модулированный сигнал, полученный на предыдущей ступени. Практическое распространение получили системы с двухкратной модуляцией ОМ-АМ, ФМ-АМ, ЧМ-АМ, ОМ-ОМ, ЧМ-ОМ, ОМ-ФМ, ОМ-ЧМ, ЧМ-ЧМ, АИМ-АМ, ФИМ-АМ, ИКМ-АМ, ИКМ-ЧМ и др. В приведенном обозначении первым указан метод модуляции на первой ступени.

Рассмотрим, какие особенности возникают при оценке потенциальной помехоустойчивости таких систем. Особенности обусловлены в основном тем, является ли модуляция последней ступени прямым или непрямым методом модуляции. Если на всех ступенях применяют прямые виды модуляции, то потенциальная помехоустойчивость определяется просто — *общий выигрыш является произведением выигрышей для каждой ступени:*

$$\mathfrak{y}=\prod_{i=1}^{n}\mathfrak{y}_i,\qquad \mathfrak{y}_1=\prod_{i=1}^{n}\mathfrak{y}_{1i}. \tag{7.77}$$

Это вызвано тем, что при прямых методах модуляции при демодуляции шум на выходе предыдущего демодулятора, который является помехой для следующей ступени демодуляции, по-прежнему является белым и полностью справедливы рассмотренные соотношения. В роли пик-фактора в системах с многоступенчатой модуляцией при анализе последующей ступени модуляции необходимо использовать пик-фактор модулированного сигнала предыдущей ступени.

Если на какой-либо одной или нескольких ступенях применяется непрямой метод модуляции, то помеха на выходе демодулятора уже не будет белым шумом, так как ее спектр определяется соотношением (7.58). В этих случаях полученные ранее аналитические соотношения требуют уточнения. Задача определения потенциальной помехоустойчивости при флуктуационной помехе с произвольным спектром сводится к рассмотренной ранее задаче с белым шумом методом, предложенным В. А. Котельниковым. Сущность этого метода заключается в том, что ко входу приемника подключают линейный четырехполюсник (выравниватель) с амплитудно-частотной характеристикой вида

$$\mathscr{K}(f)=\mathscr{K}_0/\sqrt{\Omega_2(f)}, \tag{7.78}$$

где $\mathscr{K}_0$ — постоянная. Фазовая характеристика этого четырехполюсника может быть произвольной, так как на конечные результаты она влияния не оказывает. При прохождении через этот

линейный четырехполюсник смеси сигнала и помехи каждая составляющая этой смеси преобразуется. Спектр помехи становится равномерным со спектральной плотностью

$$\Omega_0 = \mathscr{K}^2(f)\Omega_2(f) = \mathscr{K}^2_0. \qquad (7.79)$$

Если теперь при оценке помехоустойчивости приемника в качестве исходных данных использовать характеристики прошедшего через четырехполюсник сигнала и помехи со спектром $\mathscr{K}^2_0$, результаты оценки потенциальной помехоустойчивости приемника с таким выравнивающим четырехполюсником полностью совпадут с результатами решения задачи с помехой, имеющей неравномерный спектр. Поэтому выигрыш при помехе с неравномерным спектром определяется по-прежнему полученными ранее соотношениями; но в качестве сигналов рассматриваются сигналы, прошедшие выравнивающий четырехполюсник (7.78), и помеха со спектральной плотностью (7.79). Следовательно, для оценки потенциальной помехоустойчивости систем с многоступенчатой модуляцией, в которых применяют непрямые методы модуляции, первоначально решают задачу прохождения сигнала через соответствующие линейные четырехполюсники (см. § 4.4), а затем уже используют полученные ранее соотношения и общую оценку выигрыша по формулам (7.77).

Таблица 4

Система	Обобщенный выигрыш	Система	Обобщенный выигрыш
ОМ-АМ	$m^2/(m^2+\Pi^2)$	ФИМ-АМ	$0{,}6\nu^2{}_1/\Pi^2$
ФМ-АМ	$\dfrac{m^2+\nu^2{}_1}{4\Pi^2(1+m^2)}$	ОМ-ФМ	$\nu^2{}_1/4\Pi^2$
ЧМ-АМ	$\dfrac{3m^2\nu^2{}_1}{4\Pi^2(1+m^2)}$	АМ-ФМ	$\dfrac{m^2\nu^2{}_1\sqrt{2+m^2}}{8(\Pi^2+m^2)(1+m)}$
ОМ-ОМ	1	БМ-ФМ	$\nu^2{}_1/8\Pi^2$
ФМ-ОМ	$\nu^2{}_1/4\Pi^2$	АИМ-АМ	$m^2/(\Pi^2+m^2)$
ЧМ-ОМ	$3\nu^2{}_1/4\Pi^2$	ИКМ-АМ	$\dfrac{3\cdot 2^{\nu_1}}{\nu_1\Pi^2 2\ln(30F_1)}$

Данные, характеризующие обобщенный выигрыш систем с двойной модуляцией [1], сведены в табл. 4. При расчете этой таблицы учтено, что пик-фактор ФМ и ЧМ сигналов равен $\sqrt{2}$ как пик-фактор квазисинусоидального колебания, при АМ пик-фактор модулированных сигналов

$$\Pi_{АМ} = 2(1+m)/\sqrt{2+m^2}, \qquad (7.80)$$

при ОМ

$$\Pi_{ОМ} = \Pi, \qquad (7.81)$$

при балансной модуляции (БМ)

$$\Pi_{БМ} = \sqrt{2}\Pi. \qquad (7.82)$$

Сравним потенциальную помехоустойчивость системы АМ-ФМ и системы ФМ-АМ при передаче речевых сигналов. Исходные данные следующие: $m=0,8$, $\nu_1=50$, $\Pi=3$.

Для системы АМ-ФМ получим

$$\vartheta_3 = \frac{m^2 \nu_1^2 \sqrt{2+m^2}}{8(\Pi^2+m^2)(1+m)} = \frac{0,8^2 \cdot 50^2 \sqrt{2+0,8^2}}{8(3^2+0,8^2)(1+0,8)} \cong 18,7.$$

Для системы ФМ-АМ

$$\vartheta_4 = \frac{m^2 \nu_1^2}{4\Pi^2(1+m^2)} = \frac{0,8^2 \cdot 50^2}{4 \cdot 3^2 (1+0,8)^2} \cong 27.$$

Из сравнения ϑ_3 и ϑ_4 следует, что системы ФМ-АМ при передаче речевых сигналов обеспечивают в 1,44 раза больший выигрыш, чем системы АМ-ФМ. Нетрудно заметить, что это в основном обусловлено малым пик-фактором ФМ сигналов и лучшим использованием мощности передатчика. Поэтому системы ФМ-АМ обладают более высокой потенциальной помехоустойчивостью, чем системы АМ-ФМ.

Следует отметить, что порог помехоустойчивости в системах с двойной модуляцией будет наблюдаться по обеим модуляциям, если они непрямые. Если один из методов прямой, то порог определяет непрямая модуляция. Например, в системе с модуляцией ЧМ-АМ порог будет определяться ЧМ.

Контрольные вопросы

1. Что такое многоступенчатая модуляция?
2. В чем особенности оценки потенциальной помехоустойчивости систем с многоступенчатой модуляцией?
3. Как оценивают общий выигрыш для систем с прямыми методами многоступенчатой модуляции?
4. Как определяют пик-фактор сигналов в системах с многоступенчатой модуляцией?
5. Как поступают при оценке выигрыша для систем с непрямыми методами многоступенчатой модуляции?
6. Сравните потенциальную помехоустойчивость любых двух систем с двойной модуляцией.

7.7. ПОТЕНЦИАЛЬНАЯ ПОМЕХОУСТОЙЧИВОСТЬ ЦИФРОВЫХ МЕТОДОВ МОДУЛЯЦИИ

Для определенности анализ потенциальной помехоустойчивости выполним для систем с двойной модуляцией типа ИКМ-АМ. Для других случаев применения ИКМ, а также для систем с дельта-модуляцией оценку потенциальной помехоустойчивости производят аналогично.

В таблице 4 дан конечный результат оценки *обобщенного выигрыша систем ИКМ-АМ*. Покажем, как это выражение получено.

Определим через характеристики модуляции сигнала отношение сигнал/шум на выходе приемника, потом отношение сигнал/шум на входе и затем обобщенный выигрыш.

Если полезный сигнал $u(t)$ нормирован, то все значения его амплитуд лежат в интервале $[-1, 1]$. Поэтому произведение интервала квантования на число интервалов квантования равно двум:

$$\Delta uN=2. \tag{7.83}$$

При примитивном кодировании $N=m^n$. В дальнейшем будем рассматривать бинарные коды как получившие наибольшее практическое применение, следовательно, $N=2^n$. Выразим интервал квантования через параметры кода, тогда

$$\Delta u=2/N=2/2^n=1/2^{n-1}. \tag{7.84}$$

Величина Δu полностью определяет мощность шумов квантования (3.32), а мощность шумов $\mathscr{P}_2$ на выходе приемника при больших отношениях сигнал/шум на входе по существу является мощностью квантования, поэтому в соответствии с (3.32)

$$\mathscr{P}_2=\Delta u^2/3\cdot 4=1/3\cdot 2^2\cdot 2^{2(n-1)}=1/3\cdot 2^{2n}. \tag{7.85}$$

С учетом (7.50) отношение сигнал/шум на выходе приемника

$$h_2=\mathscr{P}_1/\mathscr{P}_2=1/\Pi^2\mathscr{P}_2=3\cdot 2^{2n}/\Pi^2. \tag{7.86}$$

Теперь определим отношение h_1 при ИКМ-АМ. В системе ИКМ квантованные значения следуют с частотой дискретизации $2F$, каждое квантованное значение при примитивном кодировании передается одной кодовой комбинацией длиной n. Поэтому в такой системе за одну минуту передается $2\cdot 60\,Fn$ кодовых импульсов. Предположим, что в передаче информации участвуют l ретрансляционных участков и на каждом участке вероятность ошибки при передаче одного кодового импульса равна p_0, тогда за минуту в среднем будет $120p_0lFn$ ошибочно принимаемых импульсов (ошибок). Обычно при передаче речи допускают одну ошибку в минуту (это значит, что будет слышен один щелчок в минуту), поэтому

$$p_0=1/120Fnl. \tag{7.87}$$

Допустим, что система ИКМ должна обеспечивать помехоустойчивость не хуже помехоустойчивости оптимального некогерентного приемника, тогда в соответствии с (6.85)

$$p_0=p_{22}=0{,}5e^{-h_1/2}. \tag{7.88}$$

Приравняв правые части (7.87) и (7.88), получим

$$h_1=2\ln(60Fnl). \tag{7.89}$$

Использовав (7.47), (7.96) и (7.99), найдем обобщенный выигрыш для системы ИКМ-АМ

$$\mathfrak{h}_1=3\cdot 2^{2n}/\nu_1\Pi^2 2\ln(60Fnl). \tag{7.90}$$

Обозначим длительность одного кодового импульса через τ, тогда максимальное число символов в одной кодовой комбинации

$$n^*=\Delta t/\tau, \tag{7.91}$$

где Δt — интервал дискретизации. Для передачи радиоимпульсов длительностью τ требуется, как известно, полоса $\Delta F_1 \simeq 2a/\tau$, где a — параметр, определяющий допустимые искажения формы импульсов. В отличие от систем АИМ, ФИМ и ШИМ, где в приемнике необходимо достаточно точно определять амплитуду, положение или длительность импульса, в системах ИКМ необходимо лишь зафиксировать, есть или нет импульса на входе приемника. Поэтому к форме импульса строгих требований не предъявляется (это одно из основных преимуществ ИКМ) и параметр a выбирают равным 0,5. Тогда

$$\Delta F_1=1/\tau. \tag{7.92}$$

Учтя, что $\Delta t=1/2$, $n=\Delta F_1/2F=\nu_1/2$, получим окончательное выражение для обобщенного выигрыша систем ИКМ-АМ

$$\vartheta_1=\frac{3\cdot 2^{\nu_1}}{2\nu_1 \Pi^2 \ln (30\Delta F_1)}=\frac{3\cdot 2^{\nu_1}}{2\nu_1 \Pi^2 \ln (30\nu_1 F)}, \tag{7.93}$$

где l принято равным единице.

Анализ выражений (7.90) и (7.93) позволяет сделать следующие основные выводы. С ростом коэффициента частотной избыточности выигрыш систем ИКМ-АМ экспоненциально возрастает, поэтому такие системы близки к идеальной (см. (5.89), (7.66)). При заданной полосе спектра полезного сигнала и, следовательно, фиксированном значении интервала дискретизации с ростом длины кодовой комбинации n сокращается длительность одного кодового импульса и расширяется полоса ΔF_1 канальных сигналов. Следовательно, увеличение числа N уровней квантования, с одной стороны, приводит к расширению спектра модулированных сигналов, а с другой — к уменьшению мощности помех на выходе приемника. Поэтому системы ИКМ-АМ также «обменивают» мощность на абонируемую полосу — за счет расширения полосы канальных сигналов можно уменьшить их мощность и, следовательно, обеспечить тот же выигрыш при меньших отношениях сигнал/шум на входе приемника. В отличие от систем с двойной модуляцией, где одной из ступеней является ФМ и ЧМ, которые тоже используют частотную избыточность модулированных сигналов, в системах ИКМ-АМ выигрыш возрастает не по параболическому закону, а по экспоненциальному. Следовательно, системы ИКМ-АМ лучше других осуществляют обмен мощности на полосу и намного эффективнее используют частотную избыточность. С увеличением числа уровней квантования N выигрыш систем ИКМ-АМ быстро растет. Увеличение основания кода приводит к сокращению полосы канальных сигналов, но для обеспечения одной и той же помехоустойчивости требует увеличения их мощности.

Сравним потенциальную помехоустойчивость систем ИКМ-АМ, АМ-ФМ и ФМ-АМ при передаче речевых сигналов. Возьмем исходные данные примера, приведенного в § 7.6. Выберем такое ν_1, чтобы помехоустойчивость системы ИКМ-АМ была не хуже помехоустойчивости системы ФМ-АМ. Тогда в соответствии с (7.73) и результатом решения указанного примера получим

$$\mathfrak{y}_5 = 3 \cdot 2^{\nu_1}/2\nu_1 \cdot 3^2 \ln(30\Delta F_1) \geqq 27.$$

Предположив, что $F=3400$ Гц, получим, что $\nu^*_1=15{,}2$. Так как $n=\nu^*_1/2$, то, округлив полученное значение ν^*_1 до ближайшего целого, получим $\nu_1=]\nu^*_1[=16$, где $]\cdot[$ — операция взятия большего целого. Следовательно, бинарный восьмипозиционный код обеспечивает помехоустойчивость лучшую, чем у системы ФМ-АМ. В то же время, как нетрудно заметить, полоса канальных сигналов уменьшается в 3,12 раза. Отсюда следует, что система ИКМ-АМ с числом уровней квантования $N=2^8=254$ является более помехоустойчивой и более узкополосной, чем системы ФМ-АМ и АМ-ФМ.

Как и для других методов нелинейной модуляции, для ИКМ наблюдается пороговый эффект. Его нетрудно определить из условия

$$h^*_1=2 \ln (30\Delta F_1 l). \qquad (7.94)$$

Для телефонных систем ИКМ-ЧМ $\Delta F_1=\nu_1 F=14 \cdot 3400 \simeq 47{,}6$ кГц и при $l=1$

$$h^*_1 \simeq 28{,}6. \qquad (7.95)$$

Это соответствует выигрышу $\mathfrak{y} \simeq 14$, что намного больше выигрыша других систем при пороговых значениях мощности входных сигналов. Порог в системах ИКМ обусловлен тем, что при высокой интенсивности помех выбросы помех начинают восприниматься приемником как полезные импульсы. При определенных значениях спектральной плотности помехи начинает резко расти вероятность появления ошибок при декодировании. При декодировании сказываются и отклонения характеристики фильтра нижних частот приемника от идеальной, что также увеличивает порог, но это явление проявляется незначительно.

При входных отношениях сигнал/шум, превышающих порог, пропускная способность систем ИКМ с учетом формулы (5.79) имеет вид

$$C=\Delta F_1 \log(1+h_1)=\nu_1 F \log(1+h_1)=F \log(1+h_1)^{\nu_1}. \qquad (7.96)$$

Формула (7.96) дает оценку пропускной способности «снизу», так как шум квантования имеет равномерное, а не нормальное распределение. Из (7.96) следует, что для увеличения пропускной способности систем ИКМ также целесообразно не увеличивать мощность сигналов, а расширять полосу канальных сигналов — пропускная способность экспоненциально возрастает с ростом коэффициента частотной избыточности.

Таким образом, системы с цифровыми методами передачи сигналов обеспечивают при прочих равных условиях наибольшие потенциальную помехоустойчивость и пропускную способность. «Платой» за обеспечение высоких помехоустойчивости и пропускной способности является усложнение технической реализации систем связи. Развитие микроэлектроники и вычислительной техники показывает, что этот недостаток играет все меньшую роль.

Контрольные вопросы

1. Как интервал квантования связан с параметрами кода при ИКМ?
2. Как определяют мощность шумов на выходе приемника ИКМ сигналов?
3. Чему равно отношение сигнал/шум на выходе приемника ИКМ сигналов?
4. Как определяют вероятность ошибки для систем ИКМ?
5. Как определить отношение сигнал/шум на входе приемника ИКМ сигналов?
6. Чему равен обобщенный выигрыш для системы ИКМ-АМ?
7. Как определить обобщенный выигрыш для системы ИКМ-АМ, зная коэффициент частотной избыточности, пик-фактор и полосу частот полезного сигнала?
8. Как сравнить потенциальную помехоустойчивость систем ИКМ-АМ и других систем с двойной модуляцией?
9. Как определить порог для системы ИКМ-АМ, ИКМ-ЧМ?
10. Как определить пропускную способность систем ИКМ-АМ?

7.8. ВЫВОДЫ

1. Принципиально новыми особенностями определения потенциальной помехоустойчивости передачи непрерывных сигналов являются следующие: форма и параметры передаваемых полезных сигналов обычно неизвестны; оценка помехоустойчивости зависит от выбора метрики пространства сигналов и от того, какое расстояние между переданным и принятым сигналами считается малым; основными преобразованиями при передаче непрерывных сигналов являются модуляция и демодуляция, поэтому по существу оценивают помехоустойчивость методов модуляции; как и при передаче дискретных сигналов, наиболее изучены задачи оптимального приема сигналов; основной операцией при приеме сигналов является фильтрация.

2. Сущность фильтрации заключается в том, что характеристики переданного сигнала оценивают по результатам линейного или нелинейного преобразования принятого сигнала. Различают линейную и нелинейную фильтрацию. При линейной фильтрации используют оптимальный линейный фильтр, который наилучшим образом (в среднеквадратическом смысле) выделяет модулированный сигнал из принятого, при нелинейной — автоматическую нелинейную следящую за информационным параметром систему, которая в реальном масштабе времени решает нелинейное дифференциальное уравнение фильтрации и на выходе которой получают оценку полезного сигнала. Разновидностью нелинейной фильтрации служит цифровая фильтрация, когда уравнения фильтрации решают ЦВМ. Важное достоинство цифровой фильтрации — зависимость всех характеристик фильтра только от одного пара-

метра — интервала дискретизации. При решении задач фильтрации предполагают, что статистические характеристики модулированного сигнала и помехи известны.

3. Результаты анализа оптимального приема непрерывных сигналов по критерию максимального правдоподобия позволяют полностью оценить потенциальную помехоустойчивость непрерывных систем связи. Оценка производится по обобщенному выигрышу (7.47) — частному от деления отношения сигнал/шум на выходе приемника к отношению сигнал/шум на его входе. Частное нормировано по коэффициенту частотной избыточности $\nu_1 = \Delta F_1/\Delta F$ модулированного сигнала по отношению к модулирующему. Основным результатом применения метода максимального правдоподобия являются соотношения (7.45), (7.47), позволяющие определить отношение сигнал/шум на выходе приемника и обобщенный выигрыш. Эти соотношения играют фундаментальную роль при оценке потенциальной помехоустойчивости различных методов модуляции.

4. Определение и анализ потенциальной помехоустойчивости различных методов модуляции сводится к определению и сравнению обобщенных выигрышей. Принципиальным отличием непрямых и нелинейных видов модуляции является то, что спектральная плотность шума на выходе демодулятора приемника не постоянна, а параболически зависит от частоты.

5. Пороговый эффект проявляется для всех нелинейных методов модуляции в том, что при некоторых («пороговых») отношениях сигнал/шум на входе приемника потенциальная помехоустойчивость резко падает и становится хуже помехоустойчивости однополосной модуляции, единственного линейного вида модуляции, не имеющего порога. Из-за нелинейности преобразования принятых сигналов в полезный сигнал при больших уровнях сигнала отношение сигнал/шум определяется мощностью сигналов, а при малых — мощностью помех. При некотором «пороговом» отношении сигнал/шум появляется точка перегиба на графике зависимости обобщенного выигрыша от отношения сигнал/шум на входе приемника и при уменьшении этого отношения скорость падения выигрыша резко возрастает. Для уменьшения порога применяют «сжатие» полосы пропускания приемника путем слежения за текущей шириной спектра модулированного сигнала. Уменьшение полосы приемника приводит к снижению мощности помехи на входе демодулятора, что эквивалентно относительному увеличению мощности сигнала и снижению порога.

6. Для анализа потенциальной помехоустойчивости систем с многоступенчатой модуляцией используют соотношение (7.47) для обобщенного выигрыша при однократной модуляции. Общий выигрыш определяют как произведение выигрышей для отдельных методов, входящих в многоступенчатую модуляцию. В тех случаях, когда в многоступенчатой модуляции используют непрямые методы модуляции, формулы (7.45), (7.47), (7.77) требуют уточнения. Сущность уточнения заключается в том, что задачу оценки

выигрыша при помехе с неравномерным спектром с помощью *метода выравнивающего четырехполюсника* сводят к задаче с белым шумом. В формулах применяют характеристики, полученные в результате использования этого метода. Важная особенность оценки потенциальной помехоустойчивости многоступенчатых методов модуляции заключается в том, что при анализе последующей ступени модуляции необходимо использовать пик-фактор модулированного сигнала предыдущей ступени.

7. Потенциальную помехоустойчивость цифровых методов модуляции оценивают с помощью формулы (7.47). Отношение сигнал/шум на выходе приемника выражают через пик-фактор полезного сигнала и мощность шумов квантования (7.86). Отношение сигнал/шум на входе приемника находят из условия обеспечения помехоустойчивости, равной помехоустойчивости оптимального некогерентного приемника, и выражают это отношение (7.89) через полосу спектра полезного сигнала, длину кодовой комбинации и число ретрансляционных пунктов. Обобщенный выигрыш системы ИКМ-АМ оценивают по формуле (7.93). Анализ потенциальной помехоустойчивости цифровых методов модуляции показывает, что выигрыш обеспечивается благодаря частотной избыточности канального сигнала. Из всех систем с многоступенчатой модуляцией системы цифровой передачи обладают самыми высокими помехоустойчивостью и пропускной способностью. Выигрыш и пропускная способность этих систем растут экспоненциально с увеличением коэффициента частотной избыточности, поэтому они больше всего приближаются к идеальной системе. Высокое качество этих систем достигается ценой усложнения аппаратуры.

Глава 8

КОРРЕКТИРУЮЩЕЕ КОДИРОВАНИЕ

8.1. ОСОБЕННОСТИ ПРИМЕНЕНИЯ КОРРЕКТИРУЮЩЕГО КОДИРОВАНИЯ

При корректирующем кодировании для повышения верности передачи информации воздействуют как на способ передачи, так и на способ приема. Применяют его в тех случаях, когда возможности других способов повышения верности исчерпаны. Это обусловлено усложнением систем связи при введении корректирующих устройств, ростом материальных затрат, а в ряде случаев и снижением надежности аппаратуры. Развитие корректирующего кодирования в значительной мере связано с внедрением автоматических и автоматизированных систем обработки информации, построенных на ЦВМ. Эти системы обычно являются важной составной частью иерархических систем более высокого ранга, таких, как автоматизированные системы управления воздушным движением, системы бронирования и продажи билетов, системы управления предприятиями и технологическими процессами. Для нормальной работы автоматизированных систем управления необходим обмен цифровой информацией между различными ЦВМ по телефонным и телеграфным каналам (передача данных). Допустимая вероятность ошибки при передаче одного бита информации в современных автоматизированных системах не должна превышать 10^{-6}—10^{-9}, что на 3—4 порядка меньше той, которая наблюдается в реальных каналах связи. Корректирующее кодирование направлено на согласование высоких требований к верности передачи данных и низкого качества реальных каналов, плохо приспособленных для передачи данных. Применению кодирования благоприятствует то, что большинство алгоритмов кодирования и декодирования может быть реализовано не аппаратурным, а программным способом в ЦВМ.

Возможности эффективного использования реальных каналов далеко не исчерпаны. При отношениях сигнал/шум 20—30 дБ, которые имеют место в каналах, теоретическая пропускная способность каналов при сколь угодно малой вероятности ошибок может составлять 6—10 бит/с на 1 Гц полосы канала. Следовательно, теоретическая пропускная способность телеграфного канала составляет примерно 900—1200 бит/с, телефонного — 20—30 тыс. бит/с, телевизионного — 30—50 млн. бит/с. Существующие системы позволяют получить скорость передачи информации всего

лишь 1—2 бит/с на 1 Гц полосы канала, вероятность ошибки при передаче одного бита 10^{-4}—10^{-5} и выше, если не используется корректирующее кодирование.

Для корректирования ошибок можно применять те же способы, что и для повышения скорости передачи информации. Все они направлены на увеличение объема сигнала и приближение его к объему канала. Если объем сигнала равен объему канала, то корректирования ошибок можно добиться только путем уменьшения скорости передачи информации, так как часть объема сигналов должна быть использована для корректирования. Корректирующее кодирование использует по существу все виды избыточности сигналов — временнýю, частотную и энергетическую. Если длина кодовой комбинации не фиксирована (скорость передачи информации не фиксирована), то для корректирования ошибок используют временную избыточность — кроме *информационных символов,* дополнительно вводят еще ряд символов, называемых *проверочными,* с помощью которых обнаруживают и исправляют ошибки. Эту способность кодов обнаруживать и исправлять ошибки называют *корректирующей способностью.*

Если скорость передачи информации фиксирована, ввести проверочные символы в кодовую комбинацию бинарного кода можно, лишь уменьшая длительность элементарных сигналов, что ведет к расширению их спектра. Следовательно, в этом случае корректирующее кодирование использует частотную избыточность. Чтобы отношение сигнал/шум с уменьшением длительности импульсов не падало, необходимо увеличивать амплитуду импульсов. Увеличивая амплитуду укороченного импульса, можно настолько увеличить его энергию, что вероятность ошибки при его приеме уменьшится по сравнению с вероятностью при приеме импульса неукороченной длительности. Так вводится энергетическая избыточность (корректирующая способность кода улучшается в результате повышения энергии импульсов).

Корректирующая способность кода определяется минимальным кодовым расстоянием d_0 между разрешенными кодовыми комбинациями (см. § 1.6). Максимальную кратность q_{00} обнаруживаемых ошибок определяют из следующих соображений. Если расстояние, измеренное между принятой комбинацией и какой-либо разрешенной, оказывается меньше d_0, это позволяет рассматривать принятую комбинацию как запрещенную, т. е. обнаружить ошибку. Ближайшее целое число, меньшее d_0, есть d_0—1. Поэтому кратность обнаруживаемых ошибок изменяется от 1 до d_0—1, максимальная кратность

$$q_{00}=d_0-1. \tag{8.1}$$

При *исправлении ошибок по критерию максимума правдоподобия* принимаемая комбинация отождествляется с той разрешенной, к которой она находится ближе всего. Неправильное декодирование происходит тогда, когда кодовое расстояние от принимаемой комбинации до переданной оказывается больше, чем до какой-либо

другой разрешенной. Это может случиться тогда, когда сочетание ошибок изменит более половины позиций, в которых переданная комбинация отличается от какой-либо другой разрешенной. Поэтому код с расстоянием d_0 исправляет все сочетания ошибок кратности $q_1 < d_0/2$.

Максимальная кратность полностью исправимых ошибок

$$q_{11} = \begin{cases} (d_0 - 1)/2, & \text{если } d_0 \text{ нечетно}, \\ d_0/2 - 1, & \text{если } d_0 \text{ четно}. \end{cases} \quad (8.2)$$

Для обнаружения q_0 ошибок и исправления q_1 ошибок должно выполняться неравенство

$$d_0 > q_0 + q_1. \quad (8.3)$$

Увеличение d_0 приводит к росту избыточности кода

$$r_2 = 1 - k/n = r/n, \quad (8.4)$$

где r — число проверочных символов, предназначенных для обнаружения и исправления ошибок. Наибольшей избыточностью обладают коды, обнаруживающие и исправляющие ошибки. Для них основными характеристиками являются вероятность появления ошибок на выходе декодера, минимальная избыточность, определяемая величиной d_0, эффективность кодирования (см. § 8.4).

Таблица 5

Тип канала	Вид модуляции	Скорость телеграфирования, Бод	Число переприемов по НЧ или мощность передатчика	p^*_0	α^*
Кабельный телефонный выделенный	ОФМ	1200	1	$2{,}82 \cdot 10^{-4}$	0,770
То же	ОФМ	1200	5	$1{,}29 \cdot 10^{-4}$	0,510
» »	ОФМ	1200	5	$5{,}20 \cdot 10^{-4}$	0,710
Кабельный телефонный коммутируемый (ГАТС)	ЧМ	1200	две ГАТС[1]	$2{,}00 \cdot 10^{-3}$	0,340
Радиорелейный телефонный	ОФМ	1200	3	$2{,}66 \cdot 10^{-4}$	0,606
То же	ЧМ	1200	3	$7{,}03 \cdot 10^{-4}$	0,545
Тропосферный телефонный	ОФМ	1200	—	$7{,}3 \cdot 10^{-4}$	0,439
То же	ЧМ	1200	—	$7{,}05 \cdot 10^{-4}$	0,449
Радиотелеграфный КВ	ЧМ	150	20 кВт	$2{,}85 \cdot 10^{-4}$	0,373
То же	ЧМ	50	5 кВт	$5{,}85 \cdot 10^{-3}$	0,320
» »	ЧМ	150	1 кВт	$1{,}64 \cdot 10^{-2}$	0,550

[1]) ГАТС—городская автоматическая телефонная станция.

Если ошибки независимы, то вероятность появления ошибок на выходе декодера определяют как вероятность того, что не бу-

дут исправлены ошибки кратности $q_{11}+1$ и более (см. п. 4.3.4):

$$Q_0(q_{11}+1,n)=\sum_{q=q_{11}+1}^{n} C^q{}_n p^q{}_0 (1-p_0)^{n-q}. \qquad (8.5)$$

В реальных каналах имеет место группирование ошибок, поэтому для них оценка этой вероятности [10]

$$Q_1(q_{11}+1,n)\approxeq (n/q_{11}+1)^{1-\alpha^*} p_0, \qquad (8.6)$$

где α^* — показатель группирования ошибок в канале, который оценивают экспериментально.

В таблице 5 показаны вероятности ошибок и показатели группирования для основных видов реальных каналов [10].

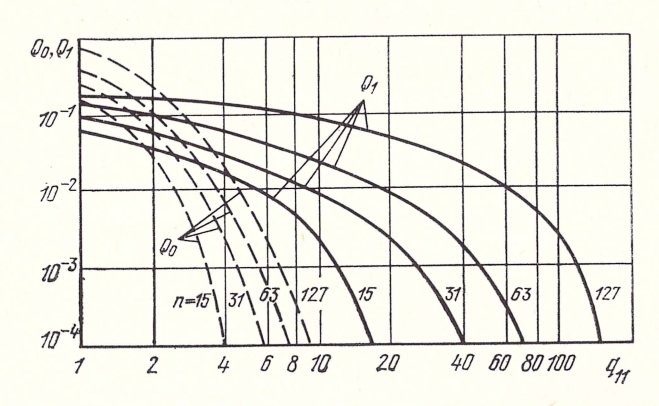


Рис. 8.1. Зависимость вероятностей некорректируемых ошибок от q_{11}

Формула (8.6) справедлива при $(q_{11}+1)/n<0,3$, что обычно соблюдается. Для сравнения данных, получаемых по формулам (8.5), (8.6), на рис. 8.1 [10] показаны зависимости $Q_0(q_{11})$ (штриховые линии) и $Q_1(q_{11})$ при различных n для радиотелеграфного канала с параметрами $p_0=1,37\cdot10^{-2}$, $\alpha^*=0,449$, $W_2=75$ Бод. Анализ графиков показывает, что формула (8.5) дает заниженные значения вероятностей появления некорректируемых ошибок из-за того, что не учитывает групповой характер ошибок. О порядке расхождения можно судить, например, по тому, что при $n=15$, $q_{11}=4$ формула (8.5) дает $Q_0=10^{-4}$, тогда как в реальном канале $Q_{11}\simeq 7\cdot10^{-2}$.

Внесение избыточности целесообразно, если применение корректирующего кодирования приводит к повышению верности. Если кратность независимых ошибок в кодовой комбинации примитивного кода равна q_0, то для гауссова канала при поэлементном приеме ортогональных сигналов целесообразные избыточность и число проверочных символов [9]

$$r_{20}<q_0(1+q_0)^{-1}, \quad r_0<nq_0(1+q_0)^{-1}. \qquad (8.7)$$

Для коррекции групповых ошибок требуется примерно в 2—3 раза меньшая избыточность. Аналогичные (8.7) соотношения более сложного характера получены с учетом группового характера ошибок и надежности корректирующих устройств. Существует информационный предел избыточности, который существенно ниже. Это объясняется тем, что избыточность вводится для наихудшего случая появления ошибок, а реально не все комбинации имеют ошибки.

С ростом длины комбинации экспоненциально возрастают объемы памяти кодера и декодера, а также задержки при кодировании и декодировании. Если использовать непосредственные способы декодирования и хранить в памяти все разрешенные комбинации, задача получения оптимальных кодов становится технически неразрешимой. Основное направление в теории корректирующего кодирования — создание таких кодов, которые не требуют хранения в памяти разрешенных комбинаций, а на основе конечного числа преобразований принятых комбинаций позволяют получать оптимальные статистические решения о том, какие комбинации передавались.

Задачи корректирующего кодирования обычно решают при следующих предположениях: избыточность эффективного кода равна нулю, кодирование выполняется двоичными сигналами, характеристики дискретного двоичного канала известны, канал является симметричным и в общем случае с памятью (см. § 4.3).

Контрольные вопросы

1. Чем обусловлено развитие корректирующего кодирования?
2. Оцените эффективность использования реальных каналов существующими системами.
3. Что такое корректирующая способность кода?
4. Как связано кодовое расстояние с кратностью обнаруживаемых и исправляемых ошибок?
5. Как определяют избыточность корректирующего кода?
6. Как определяют вероятность появления некорректируемых ошибок?
7. Как выбирают целесообразную избыточность кода?

8.2. ПРИНЦИПЫ ПОСТРОЕНИЯ КОРРЕКТИРУЮЩИХ КОДОВ

8.2.1. Классификация корректирующих кодов. Для коррекции ошибок неравномерные коды почти не применяют, поэтому в дальнейшем рассматриваются только равномерные корректирующие коды. Их общая классификация приведена на рис. 8.2. Корректирующие коды делятся на два больших класса: *блочные* и *непрерывные*. Кодовая последовательность блочных кодов состоит из отдельных кодовых комбинаций (блоков), которые кодируются и декодируются независимо. Непрерывные коды представляют непрерывную последовательность кодовых символов, ее разделение на отдельные кодовые комбинации не производится. Блочные и непрерывные коды бывают разделимые и неразделимые. В разделимых блочных кодах информационные и проверочные символы за-

нимают всегда одни и те же определенные позиции (разряды). Обозначают эти коды как (n, k)-коды, где n — длина комбинации, k — число информационных символов. *Неразделимые* коды менее распространены, к ним относится рекомендованный МККТТ стандартный телеграфный код № 3 — семиразрядный код, каждая комбинация которого содержит три единицы и четыре нуля.

Среди разделимых кодов выделяют *систематические* и *несистематические*. Систематическими кодами называют (n, k)-коды, в которых $r=n-k$ проверочных символов являются линейными

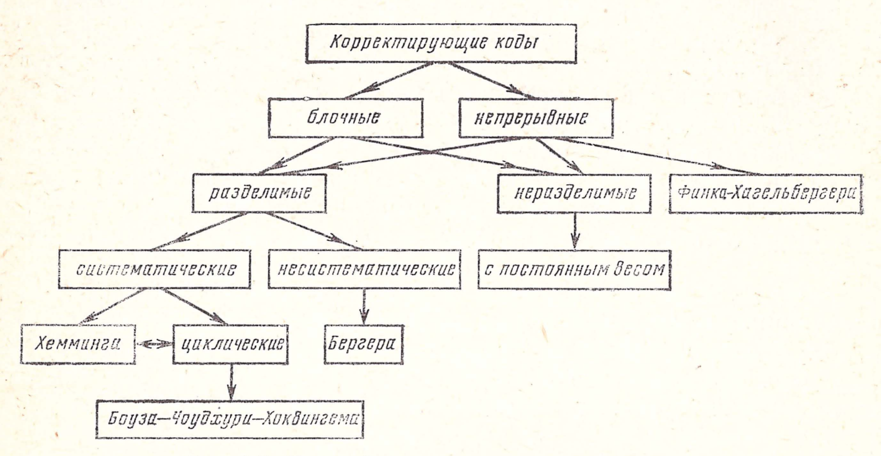


Рис. 8.2. Классификация корректирующих кодов

комбинациями информационных. Такое формирование кодовых комбинаций существенно упрощает техническую реализацию устройств *кодирования* и *декодирования* — *кодеков*. Поэтому систематические коды являются одними из наиболее распространенных. Так как новую разрешенную кодовую комбинацию можно получить линейным преобразованием двух других разрешенных комбинаций, то такие коды часто называют *линейными*. Подклассами систематических кодов являются коды Хемминга и циклические коды, например, коды Боуза—Чоудхури—Хоквингема. По установившейся традиции ряд подклассов корректирующих кодов обозначают фамилиями тех ученых, которые впервые предложили и исследовали тот или иной вид кодирования. Особенности кодов этих подклассов будут отмечены в дальнейшем.

8.2.2. О технической реализации корректирующего кодирования. Из основной теоремы Шеннона для каналов с шумами следует, что может быть обеспечена сколь угодно малая вероятность ошибочного приема при эффективности использования пропускной способности канала, близкой к единице (см. § 5.6). Для этого в симметричном канале без памяти при больших n должно выполняться условие (8.2), тогда все ошибки кратности q_{11} будут исправлены. Если непосредственно использовать условие (8.2) для технической реализации корректирующего кодирования, то необ-

ходимо применять случайный выбор разрешенных $N_0=2^k$ комбинаций из всего множества $N=2^n$ [9], это обеспечивает и любую малую вероятность появления ошибок и сколь угодно близкую к единице эффективность использования канала.

Однако практически такое случайное корректирующее кодирование неосуществимо, так как для хранения в памяти декодера всех разрешенных комбинаций при n, обеспечивающем допустимую вероятность ошибки, требуется объем памяти декодера порядка $n \cdot 2^n$ бит [9]. Например, при $n=30$ объем памяти должен быть 10^{10} бит, что намного превышает возможности современной техники. Именно поэтому все усилия исследователей направлены на разработку и создание регулярных (детерминированных) методов построения кодов, которые по свойствам были бы близки к случайному коду, но допускали более простую техническую реализацию. Примером таких регулярных кодов являются линейные систематические коды.

8.2.3. Принципы построения линейных кодов. Основным принципом построения линейных кодов является отыскание таких процедур, которые позволяют при кодировании получать все разрешенные кодовые комбинации путем конечного числа несложных линейных преобразований одной или k исходных разрешенных комбинаций, а при декодировании для обнаружения и исправления ошибок не хранить в памяти все разрешенные комбинации, а получать информацию об ошибках по результатам конечного числа относительно простых, также линейных преобразований над символами получаемых кодовых комбинаций. Основная задача оптимального построения корректирующего кода заключается в таком выборе числа разрешенных кодовых комбинаций заданной длины, чтобы обеспечивалось самое минимальное кодовое расстояние. Тогда получаемый код будет обладать наилучшей корректирующей способностью и минимальной вероятностью некорректируемых ошибок.

В линейных систематических кодах разрешенные комбинации получают путем суммирования по модулю 2 различных сочетаний исходных комбинаций. Проверочные символы получают так же, как результат суммирования по модулю 2 определенного числа информационных символов. Обнаружение ошибок основано на проверке соответствия тех проверочных символов, которые получены из принятых информационных символов, с соответствующими проверочными символами, непосредственно принятыми. Если ошибка есть, суммирование по модулю 2 проверочного символа, полученного в результате суммирования информационных символов и принятого дает единицу. Эта операция, выполненная для всех проверочных символов, даст в результате *вектор ошибок — синдром кодовой комбинации. Нулевой синдром* соответствует случаю отсутствия ошибок.

Для исправления ошибок каждому ненулевому синдрому «приписывают» наиболее вероятные для данного канала ошибки и при появлении ненулевого синдрома декодер исправляет приписанные

к нему ошибки. Если кодовая комбинация содержит r проверочных символов, то общее число ненулевых синдромов равно 2^{r-1} и таблица исправлений включает такое число строк, которое должно храниться в памяти декодера.

8.2.4. Условия формирования разрешенных кодовых комбинаций. В линейном коде сумма по модулю 2 любого конечного числа разрешенных кодовых комбинаций также является разрешенной кодовой комбинацией. Поэтому, выбрав k исходных разрешенных комбинаций, которые называют базисными, и суммируя их в i различных сочетаниях, можно построить все

$$N_0 = 2^k = \sum_{i=0}^{k} C^i{}_k \qquad (8.8)$$

разрешенных кодов комбинаций.

Чтобы каждое сочетание порождало новую разрешенную комбинацию и чтобы в конечном итоге их общее число равнялось числу, определяемому (8.8), необходимо выполнить следующие пять условий: исходные комбинации должны выбираться различными; нулевая комбинация (комбинация, состоящая из одних нулей) не должна выбираться как исходная; исходные кодовые комбинации должны быть линейно-независимы; вес каждой исходной комбинации должен быть не менее d_0 (весом называют число единиц в комбинации); кодовое расстояние между любыми парами исходных комбинаций должно быть не меньше d_0. Если k базовых комбинаций выбраны в соответствии с этими условиями, то они задают класс кодов (n, k), который включает 2^{kr} линейных кодов. Для выбора определенного кода из этого класса необходимо из множества коэффициентов γ_{ij} (см. п. 8.2.5), определяющих условия формирования проверочных символов (разрядов), выбрать те, которые минимизируют вероятность появления некорректируемых ошибок (8.6) или доставляют экстремум другому критерию оптимальности кода.

Общие методы синтеза оптимальных линейных кодов еще не созданы, но ряд оптимальных кодов при малых значениях r уже найден. К ним относятся код $(n, n-1)$, обнаруживающий одиночные ошибки, код (7.4) — код Хемминга, обнаруживающий две и исправляющей одну ошибку, код (14, 2), исправляющий ошибки до четвертой кратности включительно, и др. [1—3, 9, 10].

8.2.5. Формирование проверочных разрядов. Для линейных кодов любой проверочный разряд b_j, $j=\overline{1, r}$, формируется как линейная комбинация информационных a_i, $i=\overline{1, k}$, поэтому

$$b_j = \bigoplus\sum_{i=1}^{k} \gamma_{ij} a_i, \qquad (8.9)$$

где γ_{ij} — коэффициенты, равные 0 или 1 и выбираемые определенным образом (см. п. 8.2.4); $\bigoplus\sum$ — знак суммирования по модулю 2

(1.19). Результат операции (8.9) может быть равен 0 или 1, так как нулевая кодовая комбинация в любом линейном коде также является разрешенной.

8.2.6. Обнаружение и исправление ошибок линейными кодами. Обнаружение ошибок основано на проверке условия (8.9). Из принятых проверочных символов b'_j и тех b_j, которые получены из принятых информационных a'_i с помощью (8.9), образуют сумму

$$c_j = b'_j \oplus b_j = b'_j \oplus \sum_{i=1}^{k} \gamma_{ij} a'_i, \quad j = \overline{1,r}. \tag{8.10}$$

Если для всех j $c_j = 0$, то ошибок нет. Если хотя бы одно значение $c_j = 1$, то имеются ошибки. Вектор c_l, координатами которого служат c_j (8.10), называют синдромом l-й кодовой комбинации. Число различных синдромов, соответствующих различным сочетаниям ошибок в канале, равно числу всевозможных r-значных перестановок двоичных чисел, т. е. 2^r. При декодировании структура синдрома (число и расположение единиц) показывает, в каких разрядах произошли ошибки, и позволяет их исправить. При определенном выборе γ_{ij} синдром, рассматриваемый как двоичное число, может указывать непосредственно номер позиции (разряда) ошибки. Таблица исправления ошибок содержит $2^r - 1$ синдромов. Нулевой синдром в нее не включают, так как он указывает на отсутствие ошибок.

Процедуры формирования проверочных символов из информационных и определения синдромов с помощью линейных преобразований (8.9), (8.10), получения всех разрешенных комбинаций из k исходных также путем линейных преобразований привели к значительно более простому кодированию и декодированию, обнаружению и исправлению ошибок. Существенно сокращаются объемы памяти кодера и декодера, упрощаются кодеки, алгоритмы кодирования и декодирования.

8.2.7. Линейные коды (7.4). Эти коды называют кодами Хемминга, который впервые построил и исследовал характеристики кодов (7.4), обнаруживающих две и исправляющих одну ошибку. Выберем матрицу коэффициентов γ_{ij} в (8.9) в виде

$$(\gamma_{ij}) = \begin{pmatrix} \gamma_{15} = 1 & \gamma_{25} = 0 & \gamma_{35} = 1 & \gamma_{45} = 1 \\ \gamma_{16} = 0 & \gamma_{36} = 1 & \gamma_{36} = 1 & \gamma_{46} = 1 \\ \gamma_{17} = 1 & \gamma_{27} = 1 & \gamma_{37} = 1 & \gamma_{47} = 0 \end{pmatrix}, \tag{8.11}$$

тогда пятый, шестой и седьмой проверочные символы каждой кодовой комбинации выразятся через ее информационные символы следующим образом:

$$b_5 = a_1 \oplus a_3 \oplus a_4, \quad b_6 = a_2 \oplus a_3 \oplus a_4, \quad b_7 = a_1 \oplus a_2 \oplus a_3. \tag{8.12}$$

Код имеет следующие характеристики: $d_0 = 3$, число разрешенных комбинаций $N_0 = 2^4 = 16$, число синдромов $2^r - 1 = 7$, избыточность $r_2 = 3/7 \simeq 0{,}43$, вероятность появления некорректируемых ошибок (8.6)

$$Q_1(2,7) \approxeq (7/2)^{1-\alpha} p_0. \tag{8.13}$$

Одна из передаваемых кодовых комбинаций имеет вид 0010110, где первые три символа являются проверочными, а последующие четыре — информационными. Обратная запись комбинации показывает, что первыми принимают информационные символы для того, чтобы из них успеть образовать проверочные символы по правилу (8.9), сравнить их с принятыми проверочными по правилу (8.10) и получить синдромы для каждой кодовой комбинации.

Синдром i-й комбинации

$$c_i = (c_{3i},\ c_{2i},\ c_{1i}) = c_i\,(b_{7i} \oplus b'_{7i},\ b_{6i} \oplus b'_{6i},\ b_{5i} \oplus b'_{5i}),$$

где b_{7i} — проверочный символ, полученный из (8.9); b'_{7i} — принятый проверочный символ. Приписывая каждому синдрому номер символа, который может быть искажен с наибольшей вероятностью, получают таблицу исправлений (табл. 6). В первом синдроме не совпадают b_5 и b'_5, поэтому наиболее вероятно, что ошибочно принят проверочный символ b'_5, так как b_5, b_6, b_7 определяются через одни и те же информационные символы, а b_6 и b_7 приняты правильно. Аналогично определяют наиболее вероятные ошибки для других синдромов.

Таблица 6

Номер синдрома	Синдром	Номер ошибочного разряда кодовой комбинации
1	001	5
2	010	6
3	011	4
4	100	7
5	101	1
6	110	2
7	111	3

Таким образом, сущность *линейного корректирования ошибок* заключается в том, что каждому синдрому приписывают наиболее вероятные ошибки, по полученным кодовым комбинациям вычисляют синдромы и по номеру синдрома исправляют «приписанные» к нему ошибки.

8.2.8. Циклические коды. Существенным недостатком линейных кодов является необходимость выбора исходных разрешенных комбинаций, проверки условий формирования разрешенных комбинаций, запоминания коэффициентов γ_{ij} для формирования проверочных символов и синдромов для исправления ошибок. Поэтому поиск более простых процедур кодирования и декодирования продолжается. Он привел к появлению эффективного подкласса линейных систематических кодов — циклических кодов. Эти коды получили такое название потому, что основной операцией кодирования и декодирования является *цикл*. Цикл аппаратурно реализуют с помощью набора триггеров, объединенных в так называе-

мые регистры сдвига, и сумматоров по модулю 2. Цикл является одной из наиболее распространенных операций в вычислительной технике. Сущность циклической перестановки заключается в том, что последний символ кодовой комбинации занимает место первого, первый — второго и т. д. до тех пор, пока предпоследний символ не займет место последнего. Если циклической перестановке подвергалась разрешенная кодовая комбинация, то в результате этой операции появляется новая разрешенная комбинация.

Теория циклических кодов основана на методах высшей алгебры. (С математической точки зрения циклический код является идеалом в линейной коммутативной алгебре полинома n-го порядка по модулю x^n-1 над полем коэффициентов.) Кодовые комбинации (вектора) длиной n описывают полиномами $v_{n-1}(x)$, $n-1$ степени, в которых коэффициентами при соответствующих степенях x служат символы кодовых комбинаций. Например, комбинации 0010110 соответствует полином $x^2+x^4+x^5$ (высшие степени полинома порождаются информационными символами, которые принимают первыми).

Циклическая перестановка рассматривается как умножение $v_{n-1}(x)$ на x. Так как при сложении по модулю 2 $x^n=1$ ($x^n \oplus 1 = 0$), то замена x^n на 1 в произведении $xv_{n-1}(x)$ дает новый полином, коэффициенты которого образуют новую разрешенную комбинацию. Полином $g(x)$ степени $r=n-k$, на который делится без остатка двучлен $1+x^n$, называют *производящим (образующим) полиномом* циклического кода. Результат деления называют *проверочным полиномом* $h(x)$. Произведение

$$h(x)g(x)=1+x^n=0, \qquad (8.14)$$

поэтому полиномы $h(x)$ и $g(x)$ рассматривают как ортогональные и операцию (8.14) кладут в основу построения алгоритмов декодирования.

Кодеры и декодеры циклических кодов строят на основе регистров сдвига, охваченных обратными связями, и сумматоров по модулю 2. Разрешенные кодовые комбинации получают из одной исходной, символами которой являются коэффициенты порождающего полинома. Сначала выполняют $k-1$ циклов и получают k разрешенных базисных комбинаций. Затем линейным преобразованием этих k комбинаций получают оставшиеся 2^k-k разрешенных комбинаций. Кодер и декодер циклического кода в основном выполняют операции умножения и деления полиномов. Использование цикла для кодирования и декодирования существенно упростило аппаратурную реализацию кодеров, алгоритмов кодирования и декодирования и позволило использовать в технике связи элементную базу цифровой вычислительной техники.

8.2.9. Циклические коды (7.4). Покажем, как линейные коды (7.4) (см. п. 8.2.7) могут быть получены с помощью цикла и как строят *циклические кодеры и декодеры*. Теперь уже коэффициенты γ_{ij} выберем на основе порождающего полинома

$$g(x)=1+x+x^3. \qquad (8.15)$$

В том, что (8.15) является порождающим полиномом, нетрудно убедиться, выполнив деление x^7+1 на $g(x)$ с учетом основной особенности суммирования по модулю: $x^k+x^k=x^k-x^k=(1+1)^k=$ $=(1-1)^k=0$, тогда

$$\begin{array}{l|l} x^7+1 & x^3+x+1 \\ \cline{2-2} \underline{x^7+x^5+x^4} & x^4+x^2+x+1 \\ \quad x^5+x^4+1 & \\ \quad \underline{x^5+x^3+x^2} & \\ \qquad x^4+x^3+x^2+1 & \\ \qquad \underline{x^4+x^2+x} & \\ \qquad\quad x^3+x+1 & \\ \qquad\quad \underline{x^3+x+1} & \\ \qquad\qquad 0 & \end{array}$$

Так как деление выполнено без остатка, то $g(x)$, действительно, является порождающим полиномов кода (7, 4). Полученный полином

$$h(x)=x^4+x^2+x+1 \tag{8.16}$$

является проверочным для этого кода. Произведение

$$g(x)h(x)=1+x+x^3+x+x^2+x^4+x^2+x^3+x^5+x^4+x^5+x^7=$$
$$=1+x^7=1+1^7=0. \tag{8.17}$$

Выберем исходную кодовую комбинацию K_1 так, чтобы ее символы соответствовали значениям коэффициентов $g(x)$, тогда

$$K_1=(1101000)\cong 1+x+x^3. \tag{8.18}$$

Выполнив цикл один раз, получим

$$K_2=(0110100)\cong x+x^2+x^4,$$

второй раз —

$$K_3=(0011010)\cong x^2+x^3+x^5,$$

третий раз —

$$K_4=(0001101)\cong x^3+x^4+x^6,$$

где $\cong$ — знак соответствия двух форм записи комбинаций. Так как $k=4$, полученных комбинаций уже достаточно для образования $2^k-k=12$ остальных разрешенных комбинаций способом суммирования по модулю 2. В табл. 7 показаны все разрешенные кодовые комбинации, способ их получения и из какой комбинации можно получить выбранную, выполняя один цикл.

Важным свойством циклического кода является то, что каждая разрешенная кодовая комбинация делится без остатка на произ-

водящий полином. Это свойство используют для обнаружения и исправления ошибок. Выполнив деление принятой комбинации на производящий полином, сразу же можно выяснить, есть ошибки или нет. Если есть остаток от деления, он свидетельствует о наличии ошибок, но не указывает, какие ошибки.

Таблица 7

Комбинация	Запись		Из какой комбинации получена	Как получена комбинация
	в виде вектора	в виде полинома		
K_1	1101000	$1+x+x^3$	K_9	базисная
K_2	0110100	$x+x^2+x^4$	K_1	»
K_3	0011010	$x^2+x^3+x^5$	K_2	»
K_4	0001101	$x^3+x^4+x^6$	K_3	»
K_5	101110	$1+x^2+x^3+x^4$	K_{10}	K_1+K_2
K_6	1110010	$1+x+x^2+x^5$	K_{11}	K_1+K_3
K_7	1000110	$1+x^4+x^5$	K_4	$K_1+K_2+K_3$
K_8	0101110	$x+x^3+x^4+x^5$	K_5	K_2+K_3
K_9	1010001	$1+x^2+x^6$	K_{12}	$K_1+K_2+K_4$
K_{10}	0111001	$x+x^2+x^3+x^6$	K_6	K_2+K_4
K_{11}	1100101	$1+x+x^4+x^6$	K_{13}	K_1+K_4
K_{12}	0100011	$x+x^5+x^6$	K_7	$K_2+K_3+K_4$
K_{13}	1001011	$1+x^3+x^5+x^6$	K_{14}	$K_1+K_2+K_3+K_4$
K_{14}	0010111	$x^2+x^4+x^5+x^6$	K_8	K_3+K_4
K_{15}	1111111	$1+x+x^2+x^3+x^4+x^5+x^6$	K_{15}	$K_1+K_3+K_4$
K_{16}	0000000	0	K_{16}	K_1+K_1

Чтобы найти ошибочно принятые символы и исправить их, используют *алгоритм исправления ошибок,* который включает следующие операции:

— принятую комбинацию делят на производящий полином,

— определяют вес q остатка от деления,

— если $q \leqslant q_{11}$, то принятую комбинацию складывают по модулю 2 с полученным остатком; сумма является исправленной комбинацией;

— если $q > q_{11}$, то выполняют циклический сдвиг принятой комбинации влево на один разряд, полученную комбинацию делят на производящий полином; если вес нового остатка $q' \leqslant q_{11}$, то ее складывают с остатком, выполняют циклический сдвиг вправо на один разряд и получают исправленную комбинацию;

— если в результате первого циклического сдвига влево принятой комбинации и деления полученной комбинации на порождающий полином вес нового остатка $q' > q_{11}$, то операцию циклического сдвига и деления получаемых комбинаций повторяют до тех пор, пока вес остатка не станет равным q_{11};

— когда вес остатка стал равным q_{11}, последнюю полученную циклическими сдвигами комбинацию складывают с остатком, сумму сдвигают вправо на столько разрядов, на сколько ранее сдвигали влево принятую комбинацию, и получают исправленную комбинацию.

Кодер кода (7.4). Структурная схема кодера представлена на рис. 8.3. Триггеры (триггерные ячейки) и сумматоры по модулю 2

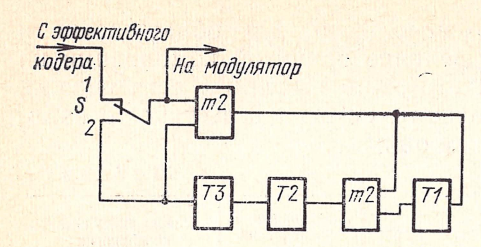


Рис. 8.3. Схема кодера кода (7.4)

обозначены в соответствии с ГОСТ 2.743—72 «Обозначения условные графические в схемах. Двоичные логические элементы».

Действие триггерной ячейки заключается в том, что при каждом воздействии на ее вход элемента кодового сигнала (0 или 1) она изменяет свое состояние на противоположное. Если в ячейке содержится символ (0) (ячейка находится в нулевом состоянии), то при появлении на ее входе символа 1 она переходит в состояние 1, а символ 0 появляется на ее выходе. Изменение состояния ячейки называют тактом или шагом. Новое состояние ячейка сохраняет до следующего шага. Сумматор по модулю 2 на каждом такте выдает суммарный сигнал 0 или 1 в зависимости от того, какие поступают входные сигналы.

Таблица 8

Номер такта	Вход регистра	Состояния триггеров регистра			Вход канала	Положение S
		T1	T2	T3		
1	1	1	1	0	1	1
2	0	0	1	1	0	
3	0	1	1	1	0	
4	1	0	1	1	1	2
5	0	0	0	1	1	
6	0	0	0	0	1	
7	0	0	0	0	0	

Рассмотрим принцип работы циклического кодера. Предположим, что с эффективного кодера поступает последовательность информационных сигналов 1001, соответствующая кодовой комбинации K_{10} (см. табл. 7). В табл. 8 показаны положение входного переключателя S на каждом такте; сигналы, поступающие на вход регистра; состояния триггеров регистра; сигналы, поступающие на модулятор; процесс формирования проверочных сигналов в регистре такт за тактом с первого по седьмой. В начальный момент времени все триггеры регистра находятся в нулевом состоянии.

Декодер кода (7.4). Структурная схема декодера представлена на рис. 8.4. Его главное назначение — обнаруживать две и исправ-

лять одну ошибку. Основная задача декодера — определить место ошибочного символа. Все остальные операции — само исправление ошибки, устранение проверочных символов и выдача на эффективный декодер информационных сигналов — являются более простыми. Схема рис. 8.4 отражает все основные особенности построения декодеров. Она включает буферный регистр, объем памяти которого равен n символов, декодирующий регистр, включающий три триггера (длиной r), и дешифратор синдромов, предназначенный для определения синдрома принимаемой комбинации и исправления ошибки.

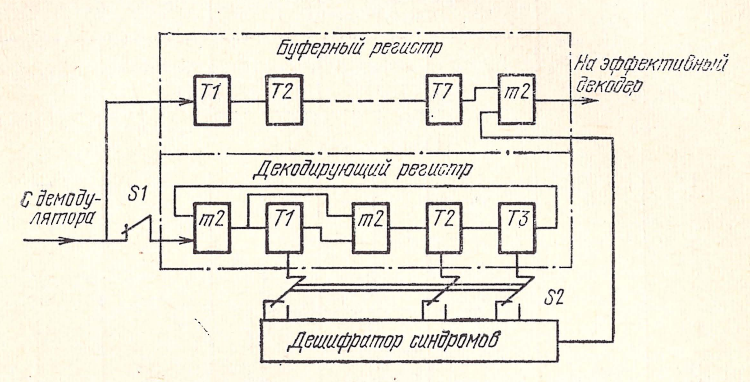


Рис. 8.4. Схема декодера кода (7.4)

Декодер работает следующим образом. Поступающие с демодулятора n символов кодовой комбинации одновременно запускают буферный и декодирующий регистры, которые вначале находятся в нулевом состоянии. После получения n-го сигнала переключатель $S1$ размыкается и подача сигналов на декорирующий регистр прекращается. На n-м такте переключатель $S2$ замыкается, присоединяя к декодирующему регистру дешифратор синдромов. Если кодовая комбинация принята правильно, то на n-м такте все триггеры декодирующего регистра находятся в нулевом состоянии, синдром является нулевым, ошибок нет, это фиксируется дешифратором синдромов и он дает разрешение информационным символам покинуть буферный регистр без коррекции. Переключатель $S2$ опять размыкается. Поступающие в буферный регистр сигналы следующей принимаемой комбинации «вытесняют» на эффективный декодер информационные сигналы предыдущей комбинации без коррекции.

Если синдром ненулевой, то хотя бы один из триггеров декодирующего регистра не находится в нулевом состоянии, что указывает на ошибку. Дешифратор синдромов обнаруживает ошибку, а также разряд ошибки и посылает в выходной сумматор буферного регистра сигнал «1» на соответствующем такте для исправления ошибки. Так как на время анализа синдрома имеет место задержка на несколько тактов, то при непрерывном поступлении

на декодер принимаемых комбинаций предусматривается их запоминание на время задержки. После завершения анализа переключатель *S1* опять замыкается и поступающие в буферный регистр сигналы «вытесняют» информационные сигналы принятой комбинации с исправлением ошибок в выходном сумматоре. Декодирующий регистр переходит в нулевое состояние, дешифратор синдромов отключается переключателем *S2*. Далее цикл повторяется.

Таблица 9

Номер такта	Сигналы на входе декодера	Состояние триггеров декодирующего регистра			Сигналы на выходе	
		*T*1	*T*2	*T*3	дешифратора	декодера
			Ошибок нет			
1	1	1	1	0	—	—
2	0	0	1	1	—	—
3	0	1	1	1	—	—
4	1	0	1	1	—	—
5	1	0	0	1	—	—
6	1	0	0	0	—	—
7	0	0	0	0	—	—
8	0	0	0	0	0	1
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	1	1	1	0	0	1
			Ошибка есть			
1	1	1	1	0	—	—
2	0	0	1	1	—	—
3	1	0	0	1	—	—
4	1	0	0	0	—	—
5	1	1	1	0	—	—
6	1	1	0	1	—	—
7	0	1	0	0	—	—
8	—				0	1
9	—	0	1	0	0	0
10	—	0	0	1	1	0
11	—	0	1	0	0	1
		1	0	0		

Для иллюстрации конкретных особенностей работы декодера рассмотрим процесс декодирования для двух случаев: когда ошибок нет и когда они есть (табл. 9). Предположим, что принятой кодовой комбинацией является комбинация K_{10} (см. табл. 7), следующей идет комбинация K_1. В табл. 9 показаны особенности декодирования для случая, когда ошибок нет. В конце 7-го такта переключатель *S1* размыкается, а *S2* — замыкается. Дешифратор определяет по нулевому состоянию всех триггеров, что ошибок нет (комбинация синдрома обведена в табл. 9 штриховой линией) и разрешает прием следующей комбинации K_1 и выдачу информационных символов принятой K_{10} на эффективный декодер.

Предположим, что в принятой комбинации K_{10} имеется ошибка в пятом разряде — третьем принятом. В табл. 9 показаны особенности декодирования в этом случае. Дешифратор синдрома на 7-м такте обнаружил ошибку. Для определения места ошибки он запускает буферный и декодирующий регистры и заставляет их выполнить еще несколько тактов. На 10-м такте, когда третий информационный символ покидает буферный регистр и поступает на выходной сумматор (сумматор исправления ошибок), на него с дешифратора поступает сигнал коррекции «1». В результате ошибка исправляется: с выходного сумматора поступает правильный сигнал «0». На время исправления ошибки сигналы комбинации K_1 запоминаются.

Декодер аппаратурно реализует общий алгоритм исправления ошибок, рассмотренный ранее. Покажем это. Разделив принятую кодовую комбинацию на образующий полином с учетом суммирования по модулю 2, получим

```
1011110 | 1011
1011    |-----
-------   1
    110
```

Остаток имеет вес $q=2$, что больше $q_{11}=1$. Поэтому выполняем циклические сдвиги принятой комбинации влево и деление получаемых комбинаций на образующий полином до тех пор, пока не получим остаток от деления с весом $q=q_{11}=1$. Таких сдвигов требуется три:

```
0111101 |1011        1111010 |1011
        |----                |----
 1011     11         1011     111
-----                -------
  1000               1000
  1011       q=3     1011           q=2
  ----               ------
   111                1110
                      1011
                      ------
                       101
```

```
1110101  | 1011
1011     |-----
-------    11
 1011
 1011        q=1
 ----
   01
```

Просуммировав последнюю кодовую комбинацию с остатком, получим 1110100. Выполнив циклический сдвиг этой комбинации вправо на три разряда, получим исправленную кодовую комбинацию 1001110.

Следовательно, для циклических кодов алгоритмы кодирования и декодирования относительно просто аппаратурно реализуются с помощью регистров и сумматоров: большие объемы памяти кодера и декодера не требуются.

8.2.10. Коды Боуза—Чоудхури—Хоквингема (коды БЧХ) — разработаны для увеличения минимального кодового расстояния и повышения корректирующей способности. Длина кодовой комбинации в них $n=2^{r+1}-1$, производящий полином находится как наименьшее общее кратное (НОК) неприводимых полиномов $\alpha_i(x)$, $i=\overline{1, m}$, где $m\leqslant d_0-2$. (Полином называют *неприводимым,* если он делится без остатка только на единицу и на самого себя. Наименьшим общим кратным совокупности неприводимых полиномов называют полином с наименьшим показателем степени, который делится на каждый из них.) Коды БЧХ обладают хорошими корректирующими способностями и позволяют обнаруживать и исправлять ошибки с учетом группирования, что очень важно для реальных каналов.

Рассмотрим код БЧХ с $d_0=5$, $n=31$ [4]. Производящий полином

$$g(x)=\text{НОК}\{\alpha_1(x)\,\alpha_3(x)\}=x^{10}+x^9+x^8+x^6+x^5+x^3+1.$$

Число избыточных символов $r\geqslant\log(n+1)\geqslant5$, избыточность $r_2\geqslant\geqslant r/n=0{,}161$, максимальная кратность обнаруживаемых ошибок $q_{00}=d_0-1=4$, полностью исправляемых ошибок $q_{11}=2$, вероятность появления некорректируемых ошибок

$$Q_1(3,\ n)\cong(n/3)^{1-\alpha}\,p_0\cong(31/3)^{1-\alpha}\,p_0. \tag{8.19}$$

Циклические коды БЧХ получили применение в аппаратуре передачи данных. Существует рекомендация МККТТ (т.41), согласно которой в среднескоростных системах передачи данных предлагается применять коды БЧХ с $d_0=4$ и $n=260$; 500; 980 разрядов. Производящий полином этих кодов

$$g(x)=x^{16}+x^{12}+x^5+1. \tag{8.20}$$

Многочисленные испытания кодов БЧХ подтвердили их высокую эффективность: при передаче данных по коммутируемым каналам телефонной сети общего пользования с $p_0>10^{-3}$ вероятность необнаруживаемых ошибок приема 8-разрядных знаков (байтов), из которых составлена информационная часть кодовой комбинации, не превышает 10^{-6}.

8.2.11. Выбор длины информационной части кодовых комбинаций при обмене информацией между ЦВМ. ЦВМ обычно обмениваются машинными словами

$$n=n_1n_2, \tag{8.21}$$

где n_1 — целые числа; $n_2=8$ разрядов — машинный слог (байт). Поэтому длина информационной части кодовой комбинации должна быть равна или кратна байту. Следовательно, число информационных сигналов

$$k=a_1 n_2, \tag{8.22}$$

где a_1 — целое число. Однако при выбранном кодовом расстоянии и числе проверочных разрядов циклические коды не дают возможности свободно выбирать длину информационной части.

Поэтому на практике прибегают к *укороченным циклическим кодам,* которые получают из полных, используя для передачи информации только те комбинации полного кода, которые содержат слева l нулей, эти нулевые сигналы в канал связи не передают. Укороченные коды обладают корректирующей способностью полных кодов, но циклический сдвиг укороченных комбинаций не всегда приводит к образованию разрешенных комбинаций. Поэтому укороченные коды часто называют *псевдоциклическими.*

8.2.12. При мажоритарном декодировании каждый информационный символ кодовой комбинации можно выразить через другие символы с помощью различных линейных соотношений и эти соотношения использовать для повышения верности. Из каждого соотношения определяют предполагаемое значение информационного символа, а окончательное решение о его значении принимается по большинству одинаковых символов для всех соотношений (методом голосования). Такая процедура легко реализуется с помощью регистров сдвига и сумматоров, она заметно повышает верность по отношению к обычному декодированию.

8.2.13. Код с постоянным весом относится к неразделимым несистематическим кодам. Разрешенными в нем являются кодовые комбинации, которые содержат определенное число единиц, одинаковое для всех комбинаций. МККТТ рекомендован код № 3 для передачи телеграфных сообщений по коротковолновым радиоканалам, которые обладают значительной асимметрией. Этот код содержит 35 разрешенных комбинаций длиной $n=7$ и весом, равным 3. Минимальное кодовое расстояние $d_0=2$, поэтому код обнаруживает все одиночные ошибки и все ошибки нечетной кратности.

В последнее время предпочтение отдается циклическим кодам, так как при использовании кодов с постоянным весом резко усложняется аппаратура и возрастает стоимость кодеков из-за неразделимости информационной и проверочной части в комбинациях.

8.2.14. В кодах Бергера, которые являются разделимыми, минимальное кодовое расстояние d_0 также равно 2.

Они также предназначены в основном для асимметричных каналов. Проверочные разряды кода Бергера представляют собой запись в двоичной форме числа единиц в информационной части кодовой комбинации. Аналогично формируются проверочные символы при декодировании. Коды Бергера обладают такой же кор-

ректирующей способностью, что и код № 3, но имеют важное достоинство — из-за разделимости кодовых комбинаций существенно упрощается построение кодеков.

8.2.15. Непрерывные коды. В простейшем непрерывном (рекуррентном) коде информационные символы a_i чередуются с проверочными b_i в определенном порядке, например

$$a_1b_1a_2b_2\ldots a_ib_i\ldots, \tag{8.23}$$

где

$$b_i = a_i \oplus a_{i+1}. \tag{8.24}$$

Коды этого класса называют кодами Финка—Хагельбергера.

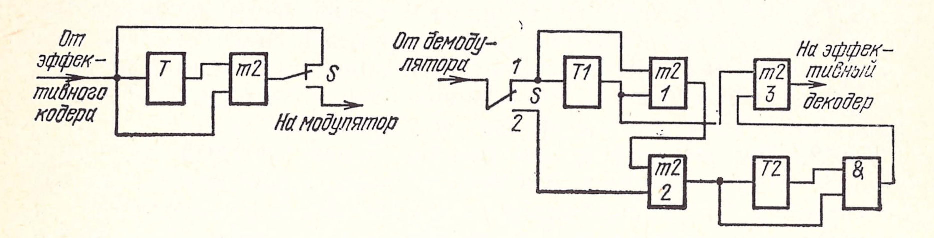


Рис. 8.5. Схема кодера непрерывного кода

Рис. 8.6. Схема декодера непрерывного кода

Рассмотрим кодирование и декодирование для такого кода. Структурная схема кодера представлена на рис. 8.5. Синхронный ключ S управляет работой кодера и чередует выдачу на модулятор информационных сигналов, которые поступают с эффективного кодера, и проверочных сигналов, которые поступают с сумматора. Триггер запоминает предыдущий информационный сигнал и выдает его на сумматор для суммирования с последующим. Структурная схема декодера представлена на рис. 8.6. Как обычно, она сложнее кодера и включает два триггера, три сумматора и одну схему совпадения. *Алгоритм работы декодера* такой: если (8.24) не выполняется для двух соседних поверочных символов, то находящийся между ними информационный сигнал следует изменить на противоположный.

Цепной код (8.23) позволяет исправить все одиночные ошибки при условии, что между любыми двумя ошибочно принятыми имеется по крайней мере три правильно принятых сигнала. Используя более совершенные коды и алгоритмы декодирования, можно исправлять групповые ошибки.

Контрольные вопросы

1. Как классифицируют корректирующие коды?
2. Какой основной принцип используется в линейных кодах?
3. Как формируют разрешенные кодовые комбинации и проверочные символы линейных кодов?
4. В чем сущность обнаружения и исправления ошибок линейными кодами?

5. Как ставится задача оптимального построения корректирующего кода?
6. Поясните работу кодера и декодера циклического кода (7,4).
7. Укажите основные особенности кодов БЧХ.
8. Как выбирают число информационных разрядов в кодовых комбинациях при обмене информацией между ЦВМ?
9. В чем сущность мажоритарного декодирования?
10. Назовите особенности кода с постоянным весом и кодов Бергера.
11. Как строят непрерывные коды, кодеры и декодеры этих кодов?

8.3. АДАПТИВНЫЕ КОРРЕКТИРУЮЩИЕ КОДЫ

Существенным недостатком многих корректирующих кодов является их слабая приспособленность к изменяющимся условиям передачи информации. Избыточность таких кодов постоянна и выбирается обычно из соображений обеспечения требуемой верности при наихудших условиях передачи. Если избыточность кода привести в соответствие с реальным состоянием канала в контролируемом интервале времени (по результатам анализа группирования ошибок), можно существенно повысить эффективность использования каналов без снижения верности. Эта идея лежит в основе построения адаптивных корректирующих кодов.

Различают методы адаптивного декодирования, когда в зависимости от числа ошибок в принимаемых кодовых комбинациях изменяют структуру или параметры алгоритмов декодирования и функции схем декодеров, и методы адаптивного кодирования, когда наряду с этим изменяют и структуру или параметры кодов, алгоритмов кодирования и схем кодеров. Функции адаптивных декодеров в значительной степени зависят от возможностей организации обратного канала, характера искажений сигналов и помех в канале, показателя группирования ошибок и других факторов. Для построения систем адаптивного кодирования требуется канал обратной связи, по которому на передающую сторону направляют информацию о качестве канала и об условиях приема.

Обнаружение ошибок адаптивными кодами систем без обратной связи позволяет обеспечить практически любую заданную верность при относительно невысокой сложности оборудования, но часть информации теряется, так как комбинации с обнаруженными ошибками потребителю не выдаются. Исправление ошибок также позволяет обеспечить требуемую верность передачи, но при отсутствии потерь информации. Платой за это является значительное увеличение длины кодовых комбинаций, до десятков тысяч разрядов, а также существенное усложнение аппаратуры. Недостатком систем без обратной связи является и то, что передатчик не получает никаких подтверждений о том, как принята информация приемником. Поэтому предъявляют очень высокие требования к надежности систем. Системы без обратной связи находят применение в случаях, когда канал обратной связи невозможно организовать или когда недопустимы задержки при передаче информации. К таким системам относятся, например, некоторые системы спутниковой связи.

Наиболее широкое применение получили системы с обратной связью (см. § 10.2), в которых повышение верности достигается обнаружением ошибок на приемном конце и повторением только неправильно принятых комбинаций. Адаптивное управление повторением информации существенно приближает избыточность кода к информационному пределу. Избыточность минимальна при отсутствии ошибок и растет с увеличением числа ошибок.

Основными задачами, которые решают при построении систем с адаптивными алгоритмами кодирования и декодирования, являются: разработка методов и аппаратуры контроля состояния каналов; оптимизация использования полученной информации о со-

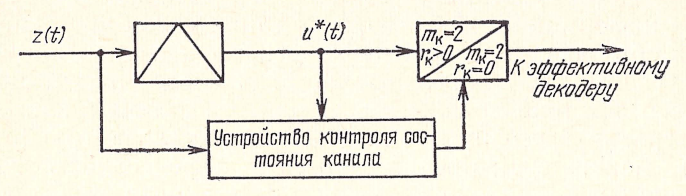


Рис. 8.7. Структурная схема адаптивного приемника

стоянии канала для изменения способа кодирования, параметров элементов сигналов кода, процедуры принятия решения и т. п.; разработка стратегий принятия оптимальных решений в приемнике в зависимости от состояния канала и характера группирования ошибок; отыскание таких алгоритмов кодирования и декодирования, при которых системы становятся инвариантными относительно статистических особенностей реальных каналов и позволяют добиться оптимальной избыточности; построение методов и аппаратуры контроля состояния канала, которые не зависят от того, какие сигналы передают в моменты контроля.

Для иллюстрации особенностей адаптивных корректирующих кодов рассмотрим простейшую структурную схему адаптивного приемника систем без канала обратной связи (рис. 8.7). Схема включает демодулятор $D_{\text{м}}$, корректирующий декодер $D_{\text{к2}}$ и устройство контроля состояния канала. На вход приемника поступают высокочастотные сигналы $z(t)$, с выхода демодулятора снимается видеопоследовательность искаженных элементарных сигналов кода $u^*(t)$, а с выхода адаптивного декодера $D_{\text{к2}}$ последовательность информационных символов в соответствии с решениями, принимаемыми декодером. Процесс адаптивного декодирования заключается в следующем. Последовательность $u^*(t)$ обрабатывается с двойной целью: для определения состояния канала и для принятия решения (для обнаружения и исправления ошибок). Если при неадаптивном декодировании решение принималось однозначно, так как каждому синдрому приписывалось одно наиболее вероятное сочетание ошибок, то при адаптивном декодировании каждому синдрому соответствует несколько вариантов сочетания ошибок в зависимости от состояния канала.

Решения о принимаемых комбинациях для одного и того же синдрома могут быть разные — они определяются состояниями канала. Если канал находится в i-м состоянии, то синдрому приписывают i-е решение. Так как состояние канала оценивают по случайному сигналу $z(t)$ и случайной последовательности $u^*(t)$, то принимаемое решение носит случайный характер. Поэтому оптимальное решение можно принять лишь с определенной вероятностью, которая в значительной мере определяется качеством работы устройства контроля канала. Если качество работы этого устройства высокое, то системы адаптивного кодирования могут обеспечить высокую эффективность передачи информации.

Теория адаптивного корректирующего кодирования интенсивно развивается, так как позволяет более полно учесть реальные условия передачи информации.

Контрольные вопросы

1. Почему разрабатывают системы адаптивного корректирующего кодирования?
2. Как классифицируют методы адаптивного корректирующего корректирования?
3. Какие особенности адаптивного кодирования имеют системы без обратной связи и с обратной связью?
4. Какие основные задачи решают в теории адаптивного кодирования?
5. Поясните сущность адаптивного декодирования.

8.4. ЭФФЕКТИВНОСТЬ КОРРЕКТИРУЮЩЕГО КОДИРОВАНИЯ

Эффективность корректирующего кодирования оценивают для различных режимов работы кода (обнаружение, исправление, обнаружение и исправление) с помощью тех или иных критериев эффективности. Наиболее распространенными из них являются: коэффициент повышения верности в той или иной форме записи, выигрыш по эквивалентной вероятности и энергетический выигрыш [1, 9, 10]. Рассмотрим, как зависит эффективность корректирующего кодирования в режиме исправления ошибок от параметров кода, характеристик канала и других условий передачи информации.

Эффективность кода в режиме исправления ошибок [10]

$$\eta_{\text{к}}=p(\geqslant 1,\ n)/Q_1(q_{11}+1,\ n), \tag{8.25}$$

где вероятность $p(\geqslant 1,\ n)$ приема кодовой комбинации с одной и более ошибками оценивают приближенно при $q/n<0{,}3$ с помощью соотношения [10]

$$p(\geqslant 1,\ n)\cong n^{1-\alpha^*}p_0, \tag{8.26}$$

а вероятность $Q_1(q_{11}+1,\ n)$ появления ошибок на выходе декодера — по формуле (8.6). Подставив (8.6) и (8.26) в (8.25), получим следующую приближенную оценку эффективности кода (оценка помечена звездочкой)

$$\eta^*_{\text{к}}\cong(q_{11}+1)^{1-\alpha^*}. \tag{8.27}$$

Следовательно, в первом приближении эффективность кода можно определить, зная показатель группирования ошибок в канале и максимальное число исправляемых кодом ошибок. Например, оценка эффективности применения кода с $q_{11}=6$ для телеграфного КВ радиоканала с $\alpha^*=0{,}55$, $n=255$ равна $\eta^*_\text{к}\simeq 7^{0,45}\simeq 2{,}4$. Следовательно, такой код повышает верность примерно в 2,4 раза. Поэтому по физическому смыслу коэффициент (8.25) можно рассматривать как «коэффициент фильтрации ошибок кодом».

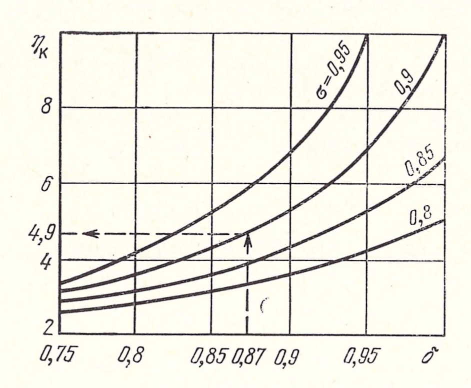


Рис. 8.8. Номограмма расчета эффективности кодирования

Введем вероятность $Q(q_{11}, n)$ исправления кодом всех ошибок кратности $1\leqslant q\leqslant q_{11}$, тогда $Q_1(q_{11}+1,\ n)=p(\geqslant 1,\ n)-Q(q_{11},\ n)$ и (8.25) примет вид

$$\eta_\text{к}=p(\geqslant 1,\ n)/[p(\geqslant 1,\ n)-Q(q_{11},\ n)]. \tag{8.28}$$

Соотношение (8.28) используют в такой форме

$$\eta_\text{к}=[1-Q(q_{11},\ n)/p(\geqslant 1,\ n)]^{-1}. \tag{8.29}$$

Дробь в знаменателе (8.29) является удельным значением вероятности исправления ошибок кодом. Для реальных кодов и каналов эта дробь изменяется в пределах 0,5—0,9 [10]. Представим эту дробь в виде

$$\frac{Q(q_{11},\ n)}{p(\geqslant 1,\ n)}=\frac{Q(q_{11},\ n)}{\sum\limits_{i=1}^{N_2} p_i}\,\frac{\sum\limits_{i=1}^{N_2} p_i}{\sum\limits_{i=1}^{N_1} p_i}=\delta\sigma, \tag{8.30}$$

где p_i — вероятность появления ненулевого синдрома с номером i (синдромы нумеруют так, чтобы вероятности p_i не возрастали с ростом i); $N_1=2^n-1$ — общее число ненулевых синдромов в кодовых комбинациях длиной n; N_2 — число ненулевых синдромов, которые наиболее вероятны для данного канала; $0\leqslant\delta\leqslant 1$ — пока-

затель согласования кода с каналом; $0 \leqslant \sigma \leqslant 1$ — показатель наиболее вероятных ошибок в канале.

Эффективность корректирующего кодирования полностью определяют параметры δ и σ:

$$\eta_к = (1 - \delta\sigma)^{-1}. \qquad (8.31)$$

В свою очередь, эти параметры зависят от параметров кода q_{11}, d_0, r, n, характеристик канала p_0, α^*, параметров элемента сигнала, вида модуляции и характеристк помех. Параметры δ и σ можно определить не только теоретически — по результатам изме-

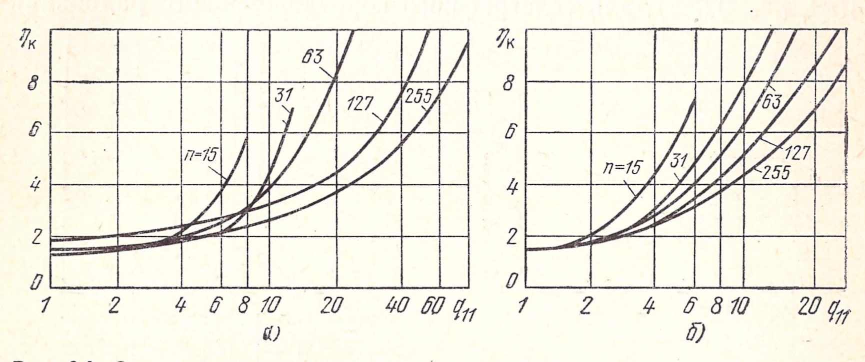


Рис. 8.9. Зависимости $\eta_к$ от q_{11} для радиорелейного телефонного канала (*а*) и для телеграфного КВ радиоканала (*б*)

рений p_0, α^* и расчетам корректирующей способности кодов, но и непосредственными экспериментальными исследованиями. Поэтому для оценки эффективности кодирования удобно пользоваться номограммой (рис. 8.8), которая позволяет по заданным δ и σ приближенно оценить эффективность $\eta_к$.

Исследуем предельное поведение $\eta_к$ в зависимости от изменения параметров δ и σ. Предел

$$\lim_{\delta \to 0,\ \sigma \to 0} \eta_к = 1. \qquad (8.32)$$

Повышение верности не может быть получено в двух случаях: если код не согласован с каналом и (или) если удельный вес наиболее вероятных ошибок пренебрежимо мал среди всех возможных ошибок.

Если код полностью согласован с каналом

$$\lim_{\delta \to 1,\ \sigma \to \sigma_0} \eta_к = (1 - \sigma_0)^{-1},$$

предельное значение эффективности полностью определяется удельным весом σ_0 наиболее вероятных ошибок в канале. Если число наиболее вероятных ошибок в канале стремится к общему числу возможных ошибок

$$\lim_{\delta \to \delta_0,\ \sigma \to 1} \eta_к = (1 - \delta_0)^{-1},$$

предельное значение эффективности определяется тем, как код согласован с каналом.

Если $\delta \to 1$ и $\sigma \to 1$, то $\eta_к \to \infty$. Это говорит о том, что, правильно выбирая код и оценивая характеристики канала, можно обеспечить сколь угодно высокую эффективность корректирующего кодирования. Пока для большинства используемых кодов и каналов $\eta_к \simeq 2-10$.

Графики зависимостей эффективности корректирующего кодирования от кратности исправляемых кодом ошибок для радиорелейного телефонного канала показаны на рис. 8.9,*а* ($p^*_0 \simeq 7{,}01 \times 10^{-4}$, $\alpha^* = 0{,}545$), для телеграфного коротковолнового радиоканала — на рис. 8.9,*б* ($p^*_0 = 1{,}31 \cdot 10^{-2}$, $\alpha^* = 0{,}449$) при различной длине кодовых комбинаций n [10]. Эффективность мала при малых значениях q_{11} и резко возрастает в области больших значений $q_{11} > 8-10$. Эффективность кодирования для телеграфного канала существенно выше. На рис. 8.10,*а*, *б* [10] для тех же каналов показаны зависимости эффективности кодирования от длины кодовых комбинаций при различной избыточности кода r_2. Эффективность корректирующего кодирования увеличивается с ростом избыточности, однако введение избыточности снижает скорость передачи информации, так как уменьшается количество информации, заключенное в одной кодовой комбинации. Избыточность кода в телеграфном канале используется более эффективно.

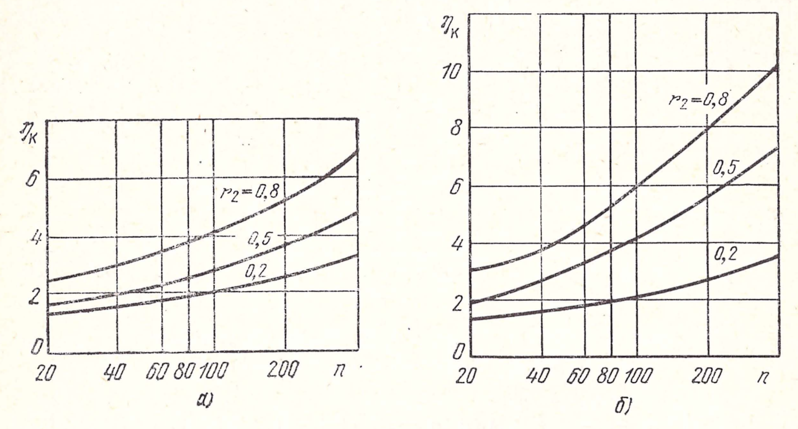


Рис. 8.10. Зависимости $\eta_к$ от n для радиорелейного телефонного канала (*а*) и для телеграфного КВ радиоканала (*б*).

Анализ графиков рис. 8.10 показывает, что существенное повышение эффективности в режиме исправления ошибок современные коды обеспечивают за счет значительного снижения скорости передачи информации. Поэтому актуальными остаются задачи построения оптимальных кодов, обеспечивающих повышение верности без существенного снижения скорости. Поэтому же широкое распространение получили системы с обратной связью и режимы обнаружения ошибок (см. § 10.2).

Контрольные вопросы

1. Как оценивают эффективность корректирующего кодирования?
2. Как связана приближенная оценка эффективности с максимальной кратностью исправляемых кодом ошибок и показателем группирования ошибок в канале?
3. Как оценивают эффективность кодирования с помощью параметра согласования кода с каналом и параметра, характеризующего удельный вес наиболее вероятных ошибок?
4. Оцените эффективность кодирования для реальных каналов.

8.5. ВЫВОДЫ

1. Развитие корректирующего кодирования во многом обусловлено внедрением различных автоматизированных систем управления, использующих обмен информацией между ЦВМ (передачу данных). Применению кодирования благоприятствует то, что большинство алгоритмов кодирования и декодирования легко реализуется программным способом в ЦВМ. Корректирующее кодирование может использовать все три вида избыточности: временну́ю, частотную и энергетическую.

Корректирующая способность кода определяется минимальным кодовым расстоянием между разрешенными кодовыми комбинациями. С ростом этого расстояния растет избыточность кода. Существуют условия, которые определяют целесообразную избыточность кода в зависимости от характера ошибок в канале. Одной из основных характеристик корректирующего кода для режима исправления ошибок является вероятность некорректируемых ошибок (8.5), (8.6).

2. Основным принципом построения корректирующих кодов является использование таких способов, которые позволяют при кодировании получать все разрешенные кодовые комбинации путем конечного обычно небольшого числа несложных преобразований одной или нескольких исходных комбинаций, а при декодировании для обнаружения и исправления ошибок не хранить в памяти все разрешенные комбинации, а получать информацию об ошибках по результатам конечного числа относительно простых преобразований над символами принимаемых комбинаций. На этом принципе построены линейные коды, циклические коды, коды Боуза—Чоудхури—Хоквингема, укороченные циклические коды, мажоритарное декодирование, непрерывные коды и др.

3. Избыточность корректирующих кодов обычно выбирают из условий обеспечения требуемой верности при наихудших условиях передачи. Если же изменять избыточность в процессе передачи в зависимости от реального состояния канала, можно существенно повысить эффективность использования канала, не снижая верности. Эта идея лежит в основе построения адаптивного корректирующего кодирования и декодирования. Адаптивные коды можно применять в системах без обратной связи и в системах с обратной связью. Использование обратной связи существенно упрощает алгоритмы кодирования и декодирования.

Основные задачи адаптивного кодирования включают разработку методов и аппаратуры контроля состояния каналов, определение стратегий принятия оптимальных решений в приемнике в зависимости от состояния канала и характера группирования ошибок, оптимизацию использования полученной информации о состоянии канала для изменения способа кодирования и декодирования и др.

4. Для оценки эффективности корректирующего кодирования в режимах обнаружения ошибок, исправления ошибок, обнаружения и исправления ошибок используют различные критерии. Удобными критериями для оценки эффективности кода в режиме исправления ошибок являются коэффициенты повышения верности (8.25), (8.27), (8.28), (8.31).

Эффективность корректирующего кодирования определяется двумя основными факторами: согласованием кода с каналом и удельным весом наиболее вероятных ошибок в канале. Эффективность кодирования для реальных каналов мала при небольших кратностях исправляемых ошибок и резко возрастает в области $q_{11}>8—10$. Современные коды обеспечивают высокую эффективность в основном в результате увеличения временной избыточности и снижения скорости передачи информации. Поэтому актуальными являются задачи построения кодов, обеспечивающих существенное повышение верности без значительного снижения скорости передачи информации.

Глава 9

УПЛОТНЕНИЕ ЛИНИЙ СВЯЗИ И ИНФОРМАЦИОННЫЕ ПОТОКИ В СЕТЯХ

9.1. ЭЛЕМЕНТЫ ТЕОРИИ РАЗДЕЛЕНИЯ СИГНАЛОВ

Во введении и в § 1.7 отмечалось, что основной проблемой уплотнения линий связи является проблема разделения канальных сигналов. Рассмотрим элементы линейной теории разделения сигналов, начало которой было положено Д. В. Агеевым в 1935 г. Обозначим i-канальный сигнал в виде

$$s_i(t)=a_i x_i(t), \tag{9.1}$$

где a_i — коэффициент, отображающий передаваемое сообщение; $x_i(t)$ — сигнал-переносчик (так как $a_i(t)$ является «медленной» функцией по сравнению с $x(t)$, то на интервале существования $x(t)$ можно считать, что $a_i(t) \simeq a_i$). Линейный передаваемый сигнал

$$s(t)=\sum_{i=1}^{n} a_i x_i(t), \tag{9.2}$$

где n — число каналов. Этот сигнал в процессе передачи по линии связи искажается из-за отклонения характеристик линии от идеальных и из-за воздействия помехи $\xi(t)$, поэтому сигнал на входе приемника многоканальной системы

$$z(t)=\sum_{i=1}^{n} a_i x_i(t)+\xi(t). \tag{9.3}$$

Отличительной особенностью работы многоканальных систем является наличие переходных помех $\xi_{ji}(t)$, обусловленных прохождением части мощности i-го канала в j-й канал из-за несовершенства аппаратуры уплотнения и разделения сигналов. Эти помехи описывают обычно одной общей флуктуационной помехой $\xi(t)$ (9.3).

Выполним оптимальную линейную обработку (6.1) принятого сигнала с учетом (6.4), тогда

$$\int_0^T z(t)\, x_i(t)\, dt=\int_0^T \sum_{i=1}^{n} a_i x^2{}_i(t)\, dt+\int_0^T \xi(t)\, x_i(t)\, dt. \tag{9.4}$$

Если сигналы-переносчики образуют ортонормированную систему функций (2.4), то

$$\int_0^T z(t)\, x_i(t)\, dt=a_i+\int_0^T \xi(t)\, x_i(t)\, dt. \tag{9.5}$$

Из (9.5) следует, что применение ортонормированных сигналов-переносчиков и линейной оптимальной обработки сигналов позволяет выделить полезный сигнал a_i с максимальным отношением сигнал/шум (6.9). Мощность шума существенно зависит от способа уплотнения сигналов.

Условие ортонормированности сигналов-переносчиков является достаточным, но не необходимым. Необходимым условием разделимости сигналов является менее строгое условие линейной независимости сигналов переносчиков

$$\sum_{i=1}^{n} a_i x_i (t) \equiv 0 \qquad (9.6)$$

только тогда, когда все a_i одновременно равны нулю. Физический смысл этого условия определяется соотношением (9.3). Если бы условие (9.6) не выполнялось, то принятый сигнал мог содержать только помеху из-за того, что полезные сигналы «взаимно компенсировались». Это могло случиться тогда, когда подобрались бы такие, не равные нулю коэффициенты a_i, при которых условие (9.6) удовлетворяется.

Анализ (9.5) показывает, что приемник многоканальной системы должен выполнять оптимальную корреляционную обработку принятого линейного сигнала с использованием известных сигналов-переносчиков. Это наиболее общий способ разделения сигналов — разделение сигналов по форме. Частными случаями этого способа являются частотное, временно́е, фазовое, кодовое разделение и др. [1—3, 8].

Контрольные вопросы

1. Как в теории линейного разделения канальных сигналов представляют передаваемый и принятый линейные сигналы?
2. Какая обработка принятого линейного сигнала является оптимальной?
3. Сформулируйте и поясните необходимое условие разделимости сигналов.
4. Какие частные случаи разделения сигналов по форме используют в многоканальной связи?

9.2. ПРИНЦИПЫ УПЛОТНЕНИЯ ЛИНИЙ СВЯЗИ [1—3, 8]

9.2.1. Частотное разделение канальных сигналов. При частотном разделении канальных сигналов переносчики имеют различные частоты. Поэтому модулированные (канальные) сигналы занимают неперекрывающие полосы частот в общей полосе линии связи и являются ортогональными. Выделение канальных сигналов в приемнике производится с помощью полосовых фильтров. Центральные частоты фильтров соответствуют частотам переносчиков, а полосы прозрачности — ширине спектров модулированных сигналов.

Для эффективного разделения канальных сигналов необходимо, чтобы выполнялись два условия: частоты сигналов-переносчиков должны быть различны и разнесены на интервал, больший или равный ширине спектра модулированного сигнала; отклонение реальных характеристик полосовых фильтров от идеальных не должно влиять на качество разделения, поэтому необходимо оставлять защитный интервал частот между каналами.

Для модуляции канальных переносчиков можно применять все известные способы. Ясно, что более экономное использование полосы частот линии связи имеет место при однополосной модуляции, так как в этом случае ширина спектра модулированного сигнала равна ширине спектра полезного сигнала.

В технике радиосвязи применяют двойную модуляцию. Полученным в результате уплотнения линии связи групповым сигналом модулируют основную несущую, а канальные частоты называют поднесущими. В приемнике сначала выполняют демодуляцию с целью получения линейного сигнала, а затем уже производится разделение канальных сигналов. Системы с поднесущими нашли широкое применение в радиорелейной связи и в телеметрии. Качество передачи информации в таких системах определяется обоими видами модуляции. При оценке помехоустойчивости двойной модуляции необходимо учитывать и переходные помехи.

Исследования показали, что система ОМ-ЧМ, получившая распространение в радиорелейных УКВ линиях связи, обладает сравнительно высокой помехоустойчивостью и относительно неплохой эффективностью использования полосы

частот линии. Недостатками этой системы являются снижение эффективности использования полосы частот линии при увеличении индекса модуляции, наличие порога помехоустойчивости и появление переходных помех при многолучевом распространении радиоволн. Большое внимание уделяется системе ОМ—ОМ, которая лишена указанных недостатков. Ее недостатком является усложнение аппаратуры формирования группового и линейного сигналов, а также аппаратуры выделения полезных сигналов. Для многоканальной телеграфной связи используют и диапазон коротких волн, обычно применяют системы ЧМ-ОМ, а также ОФМ-ОМ, если число каналов невелико. В радиорелейных линиях связи для передачи телеграфной информации выделяют часть телефонных каналов, которые вторично уплотняют телеграфными каналами.

Основные достоинства частотного разделения: простота технической реализации, высокая помехоустойчивость, возможность организации любого числа каналов, а также последующего его увеличения, если позволяет общая полоса пропускания линии связи. Недостатки: неизбежное расширение полосы используемых системой частот при увеличении числа каналов, относительно низкая эффективность использования полосы частот линии из-за потерь на «расфильтровку», громоздкость и высокая стоимость аппаратуры, обусловленные в основном большим количеством фильтров.

9.2.2. Временно́е разделение канальных сигналов. Многоканальные системы с временным разделением канальных сигналов широко применяют для передачи как аналоговой, так и дискретной информации. Сущность временного разделения заключается в том, что все каналы поочередно используют одну и ту же полосу частот линии связи. Временно́е разделение может применяться только при импульсной модуляции (см. § 3.5), когда из-за большой скважности между импульсами одного канала образуется большой интервал времени, в котором можно разместить импульсы других каналов.

Передающее и приемное устройства многоканальной системы с временным разделением сигналов имеют электронные переключатели, назначение которых периодически и синхронно подключать к линии передатчик и приемник каналов системы. Частоту F повторения канальных сигналов (величина, обратная периоду подключения к линии одного канала) выбирают согласно теореме Котельникова. С учетом реальных характеристик устройств обычно

$$F=(2{,}5-5)F_{\text{макс}}, \tag{9.7}$$

где $F_{\text{макс}}$ — верхняя граничная частота полезных сигналов. Для синхронизации работы переключателей передают вспомогательные синхронизирующие импульсы, для которых отводят один или несколько каналов.

При временном разделении используют различные виды импульсной модуляции: фазоимпульсную, широтно-импульсную, импульсно-кодовую, дельта-модуляцию и др. При использовании радиолиний применяют двойную модуляцию, например ФИМ-АМ, ФИМ-ЧМ, ИКМ-ЧМ, ИКМ-ОФМ.

В системах с временным разделением отсутствуют канальные полосовые фильтры, стоимость которых достигает 40% стоимости систем с частотным разделением. При временном разделении каналов просто осуществить выделение каналов на промежуточных станциях, без каких-либо ограничений количества выделяемых каналов и с сохранением высокого качества передачи. Аппаратура этих систем имеет малые габариты и массу, что обусловлено широким применением микромодулей, интегральных схем, элементов цифровой вычислительной техники.

Основной недостаток систем с временным разделением — необходимость обеспечения синхронной работы коммутаторов каналов передатчика и приемника, относительная сложность изготовления аппаратуры на большое количество каналов, рост полосы занимаемых частот с увеличением числа каналов, что обусловлено уменьшением длительности передаваемых канальных импульсов. Расширение полосы по сравнению с системами с частотным разделением оправдывается высокой помехоустойчивостью импульсных методов передачи (ФИМ, ИКД, ДМ), что является существенным преимуществом систем с временным разделением.

Преимущества систем с временным разделением каналов обусловили их широкое распространение в телефонии, особенно систем с импульсно-кодовой моду-

ляцией и временным делением каналов (ИКМ-ВД). В этих системах применяют компандирование речевых сигналов (см. § 3.5). В результате общее количество уровней квантования, необходимое для обеспечения высокого качества, резко уменьшается. Отличное качество передачи речевой информации при неравномерном квантовании достигается при $N_1=2^7=128$ уровнях, хорошее — при 64 и удовлетворительное — при 32. В современных системах используют 8-разрядный код — семь информационных импульсов и восьмой для передачи сигналов управления работой автоматических телефонных станций.

Системы ИКМ-ВД находят применение для уплотнения как городских, так и междугородных телефонных линий. Их используют для передачи речевых, телевизионных, видеотелефонных сигналов и высокоскоростной передачи данных. Восьмизначный код имеет кодовые комбинации, длина которых соответствует одному машинному слогу (байту). Поэтому системы ИКД-ВД хорошо сопрягаются с входными и выходными устройствами ЦВМ. Непосредственный синхронный ввод дискретной информации в ИКМ каналы позволяет получить скорость передачи в стандартном телефонном канале около 64 кБод. При обычном вводе информации, когда применяют ее двойное преобразование (аналого-цифровое и цифроаналоговое), максимальная скорость всего лишь порядка 10 кБод.

9.2.3. Частотно-временное разделение сигналов применяют для того, чтобы использовать преимущества как частотного, так и временного уплотнения. Системы с частотно-временным разделением строят по следующему принципу: предварительно производится уплотнение по времени, затем образовавшиеся группы каналов подают на вход системы с частотным уплотнением, в которой каждая группа каналов работает на своем несущем колебании. При таком комбинированном методе уплотнения линий связи значительно увеличивается число каналов системы и существенно упрощается аппаратура по сравнению с системами только частотного уплотнения.

9.2.4. Разделение канальных сигналов по форме лежит в основе построения широкополосных систем связи с шумоподобными сигналами-переносчиками (см. § 3.3). Шумоподобные сигналы различной формы удовлетворяют условию (9.6) линейной независимости. Для передачи можно использовать одну и ту же полосу частот линий связи и передачу производить в одни и те же интервалы времени. Поэтому эффективность использования линий связи существенно возрастает. Для разделения канальных сигналов в приемнике наиболее часто используют квазиортогональность (см. § 3.3) шумоподобных сигналов.

9.2.5. Кодово-адресное разделение канальных сигналов объединяет преимущества частотного и временного разделения и разделения сигналов по форме. В системах с кодово-адресным разделением, чаще всего использующим УКВ радиолинии, каждому абоненту (станции) присваивается адрес, который представляет определенную кодовую последовательность, известную всем абонентам системы. Абоненты одновременно используют общую полосу частот линии связи, поэтому системы с кодово-адресным разделением называют системами со свободным доступом. Связь между любыми двумя абонентами может осуществляться в любое время и вне зависимости от использовании линии в это время другими абонентами. Основным фактором, сдерживающим рост числа одновременно работающих абонентов, является увеличение уровня взаимных помех.

В качестве сигналов-переносчиков используют ансамбль сигналов, обладающих хорошими корреляционными свойствами и прежде всего малыми значениями взаимокорреляционной функции. Такие сигналы, как уже отмечалось (см. § 3.3), называют квазиортогональными. В системах с кодово-адресным разделением используют как непрерывные, так и дискретные сигналы. В системах с непрерывными сигналами используют шумоподобные сигналы или двоичные псевдослучайные последовательности (см. § 3.3).

Во многих системах с кодово-адресным уплотнением в качестве сигналов используют набор элементов частотно-временной матрицы. С помощью ФИМ или ДМ непрерывные сигналы преобразуются в последовательность импульсов определенной длительности. Формирующее устройство передатчика (шифратор) кодирует каждый двоичный информационный символ в дискретный адрес — набор нескольких (трех-четырех) импульсов с различной частотой заполнения. Управляемыми параметрами шифратора являются временные интервалы между импульсами и частоты заполнения видеоимпульсов адресной последовательности.

Эти параметры и играют роль параметров селекции при разделении канальных сигналов.

Общее число время-частотных комбинаций импульсов в адресных последовательностях достигает нескольких десятков тысяч (общее число абонентов). В приемнике перед демодулятором стоит дешифратор адресной последовательности, который выделяет информационные сигналы только своего адреса. Демодулятор преобразует выделенные информационные импульсы в полезные сигналы. Приемник в таких системах работает как цифровой согласованный фильтр.

Важными достоинствами систем с кодово-адресным уплотнением линий связи являются высокая эффективность и надежность связи, более низкая стоимость аппаратуры по сравнению с системами частотного разделения, высокая помехоустойчивость.

Контрольные вопросы

1. В чем сущность частотного, временного, частотно-временного, кодово-адресного разделения и разделения канальных сигналов по форме?
2. Укажите достоинства и недостатки наиболее распространенных способов разделения канальных сигналов.

9.3. ПРОПУСКНАЯ СПОСОБНОСТЬ УПЛОТНЕННЫХ ЛИНИЙ СВЯЗИ

Идеальное разделение канальных сигналов практически невозможно ни при каком способе разделения. Поэтому реально всегда имеют место переходные помехи — на выходе i-го канала появляются те или иные составляющие сигналов j-го канала ($j \neq i$, $j = \overline{1, n}$). Появление переходных помех обусловлено погрешностями формирования канальных сигналов, искажениями сигналов в линии связи и несовершенством разделяющих устройств приемника. Переходные помехи независимо складываются и при большом числе каналов образуют гауссовский флуктуационный шум. Следовательно, при расчете пропускной способности уплотненных линий связи (многоканальных линий связи) необходимо, кроме обычной аддитивной помехи, учитывать и переходные помехи.

Реальные полезные сигналы имеют или неограниченную полосу спектра при конечной длительности, или неограниченную длительность при ограниченной полосе. Это приводит к тому, что и модулированные (канальные) сигналы обладают этими же свойствами. Следовательно, всегда будут погрешности формирования канальных сигналов из-за одновременного ограничения и полосы спектра, и длительности сигналов. Из-за нелинейных искажений сигналов в линии связи появляются побочные составляющие, которые, попадая на вход каналов вместе с полезными составляющими, играют роль аддитивного шума. Из-за несовершенства разделения канальных сигналов на демодуляторы отдельных каналов «просачиваются» составляющие других сигналов, которые также играют роль шума.

Обозначим через λ_{ij} коэффициент взаимного мешающего действия j-го канала на i-й канал. Этот коэффициент показывает, какая часть мощности сигнала j-го канала «просачивается» в i-й канал. Средняя мощность $\mathscr{P}_i$ переходных помех для i-го канала от j-го канала

$$\mathscr{P}_i = \lambda_{ij} \mathscr{P}_{1j}, \tag{9.8}$$

где $\mathscr{P}_{1j}$ — средняя мощность j-го канального сигнала.

Общая мощность помехи в i-м канале

$$\mathscr{P}_{2i} = M\,[\xi^2{}_i\,(t)] + \sum_{j=1}^{n} \lambda_{ij} \mathscr{P}_{1j}, \tag{9.9}$$

где $\xi_i(t)$ — аддитивная помеха в i-м канале. Величину $\mathscr{P}_{2i}$ и необходимо использовать при оценке пропускной способности уплотненной линии связи. Подставив (9.9) в формулу Шеннона (5.79) для гауссова канала, получим

$$C_n = \sum_{i=1}^{n} \Delta F_i \log \left[1 + \frac{\mathscr{P}_{1i}}{\mathscr{P}'_{2i} + \sum_{j=1}^{n} \lambda_{ij} \mathscr{P}_{1j}} \right], \tag{9.10}$$

где ΔF_i — полоса частот i-го канала; $\mathscr{P}'_{2i} = M[\xi^2{}_i(t)]$ — мощность флуктуационной помехи в i-м канале; $\mathscr{P}_{1i}$ — средняя мощность i-го канального сигнала.

Из общей формулы (9.10) можно получить конкретные формулы для различных способов уплотнения. Например, если $\lambda_{ij} = \lambda$, $\mathscr{P}_{1j} = \mathscr{P}_{1i} = \mathscr{P}_1$, $\mathscr{P}'_{2i} = \mathscr{P}'_2$,

$$\frac{\mathscr{P}'_{2i}}{\sum_{j=1}^{n} \lambda_{ij} \mathscr{P}_{1j}} = \frac{\mathscr{P}'_2}{n\lambda \mathscr{P}_1} \ll 1, \tag{9.11}$$

то

$$C_n = \sum_{i=1}^{n} \Delta F_i \log \left(1 + \frac{1}{n\lambda} \right). \tag{9.12}$$

Условие (9.11) означает, что суммарная мощность переходных помех намного превосходит мощность флуктуационной помехи. Если взаимное влияние каналов одинаково, мощности канальных сигналов и помех в различных каналах одинаковы, то нельзя увеличить пропускную способность уплотненной линии связи, увеличивая среднюю мощность сигналов — необходимо или уменьшать взаимное влияние каналов, чтобы $\lambda \to 0$, или идти на уменьшение числа каналов n.

Уплотнение линий связи позволяет существенно повысить эффективность их использования. Создание линий связи, особенно кабельных, связано со значительными расходами на сооружение сетей, поэтому развитие методов уплотнения целесообразно и с технической, и с экономической точки зрения.

Контрольные вопросы

1. Назовите основные причины появления переходных помех?
2. Как рассчитать пропускную способность уплотненной линии связи?

9.4. ОПТИМИЗАЦИЯ ПРОПУСКНОЙ СПОСОБНОСТИ УПЛОТНЕННЫХ ЛИНИЙ В СЕТЯХ СВЯЗИ

Оптимизация пропускной способности уплотненных линий в сетях связи является актуальной проблемой. Решить ее можно с помощью теории сетей связи.

Покажем на простейшем примере, как можно оптимизировать пропускную способность многоканальных линий связи (МЛС) сетей с коммутацией сообщений и фиксированной процедурой выбора пути. Задачу оптимизации необходимо рассматривать как технико-экономическую. В качестве критерия оптимальности в прямой постановке задачи выберем среднее время T задержки сообщений в сети (см. § 10.4), в качестве ограничения — арендную плату $\mathscr{G}$ за пользование сетью. Необходимо так выбрать пропускную способность МЛС, чтобы при заданной арендной плате обеспечить минимальное среднее время задержки сообщений. В обратной постановке задачи в качестве критерия оптимальности выступает $\mathscr{G}$, а в качестве ограничения T.

Строгое решение задачи вызывает большие трудности. Поэтому примем следующие допущения: потоки сообщений являются стационарными и пуассоновскими, сообщения — одноадресными, законы распределения времени передачи сообщений во всех узлах связи — показательными, длина сообщений на входе любого внутреннего узла связи — независима и случайна, емкость накопителей сообщений на узлах — бесконечна. Предположим, что помехи и отказы аппаратуры линий и узлов связи отсутствуют. Эти допущения позволяют получить результаты в первом приближении, полезные для практики.

Введем следующие обозначения: Λ_i — интенсивность потока сообщений на входе i-й линии связи (сообщ./сут.); $\mathscr{M}_i^{-1}$ — средняя длина сообщений, поступающих в i-ю линию связи (слов/сообщ.); C — информационная пропускная способность одноканальной линии связи (слов/сут); C_i — эквивалентная информационная пропускная способность i-й многоканальной линии связи, рассматриваемой в определенном смысле как одноканальной (слов/сут); m — количество узлов связи в информационной сети; $\mathscr{M}_i C$ — интенсивность передачи сообщений по одноканальной линиии связи (сообщ./сут); $\mathscr{M}_i C_i$ — эквивалентная интенсивность передачи сообщений по многоканальной линии связи (сообщ./сут); n_i — число каналов в i-й многоканальной линии связи; s_i — арендная плата в единицу времени за использование единицы пропускной способности i-й линии связи (руб./слово); $\mathscr{G}$ — арендная плата в единицу времени за пользование информационной сетью (руб/сут·сеть); γ_{ij} — интенсивность поступления сообщений в сеть от источника в узле i к адресату в узле j (сообщ./сут); $\gamma=\sum\limits_i\sum\limits_j \gamma_{ij}$ — общая интенсивность поступления сообщений от внешних источников (сообщений/сут), определяемая суммированием γ_{ij} по всем источникам и адресатам; $\mathscr{M}^{-1}$ — средняя длина сообщений по всем источникам; C_0 — суммарная емкость всех линий связи в сети; $\rho=\gamma/\mathscr{M}C_0$ — нагрузка сети, для установившегося режима $\rho<1$.

Для строгого определения T следует рассматривать многоканальную сетевую систему массового обслуживания с ожиданием. Целесообразен более простой путь замены сетевой многоканальной системы некоторой эквивалентной по среднему времени задержки сообщений сетевой одноканальной системой. Если поток сообщений в i-й многоканальной системе распределяется равномерно между n_i каналами, то приближенно эквивалентная одноканальная i-я система имеет интенсивность передачи сообщений

$$\mathscr{M}_i C_i=\Lambda_i+\mathscr{M}_i C-\Lambda_i/n_i. \tag{9.13}$$

Отсюда

$$n_i=\Lambda_i[\Lambda_i+\mathscr{M}_i(C-C_i)]. \tag{9.14}$$

Среднее время задержки сообщений в сети

$$T=\sum_{i=1}^{m} T_i\Lambda_i/\gamma,$$

где Λ_i/γ — относительное количество сообщений, которые задерживаются в i-м узле; T_i — задержка сообщений в i-м узле. Учтем, что $T_i=1/(\mathscr{M}_i C_i-\Lambda_i)$, тогда

$$T=\sum_{i=1}^{m}\frac{\Lambda_i}{\gamma\,(\mathscr{M}_i C_i-\Lambda_i)}\ \text{[сут/сообщ.]}. \tag{9.15}$$

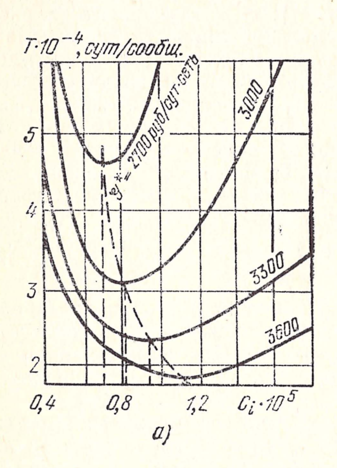
Арендная плата за пользование сетью

$$\mathscr{G}=\sum_{i=1}^{m} s_i C_i\ \text{[руб./сут·сеть]}. \tag{9.16}$$

Оптимальную пропускную способность МЛС определим методом неопределенных множителей Лагранжа. Для прямой задачи вспомогательная функция Лагранжа имеет вид

$$L(C_i, \lambda) = \sum_{i=1}^{m} \frac{\Lambda_i}{\gamma(\mathcal{M}_i C_i - \Lambda_i)} + \lambda\left(\mathcal{G}^* - \sum_{i=1}^{m} s_i C_i\right), \tag{9.17}$$

где λ — множитель Лагранжа; $\mathcal{G}^*$ — допустимая арендная плата. Продифференцировав (9.17) по C_i и λ, получим систему $m+1$ уравнений относительно неизвестных C_i $(i=\overline{1, m})$ и λ.



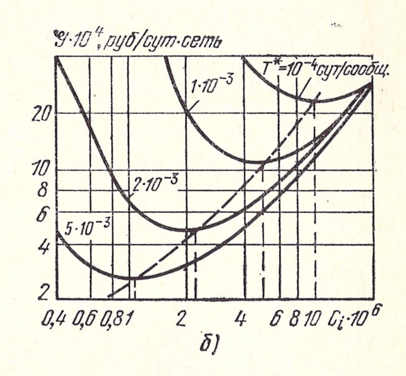


Рис. 9.1. Зависимости T от c_i (a) и $\mathcal{G}$ от c_i ($б$)

Решив эти уравнения, получим оптимальную пропускную способность МЛС, при которой среднее время задержки сообщений минимально,

$$C_{i\text{опт}} = \frac{\Lambda_i}{\mathcal{M}_i} + \sqrt{\frac{\Lambda_i}{\mathcal{M}_i C_i}} \left(\mathcal{G}^* - \sum_{j=1}^{m} \frac{\Lambda_i s_j}{\mathcal{M}_j}\right) \Bigg/ \sum_{j=1}^{m} \sqrt{\frac{\Delta_i s_j}{\mathcal{M}_j}}, \tag{9.18}$$

$$T_{\text{мин}} = \left(\sum_{i=1}^{m} \sqrt{\frac{\Lambda_i s_i}{\mathcal{M}_i}}\right)^2 \left[\gamma\left(\mathcal{G}^* - \sum_{j=1}^{m} \frac{\Lambda_i s_j}{\mathcal{M}_j}\right)\right]^{-1}. \tag{9.19}$$

Аналогично получаем оптимальное решение для обратной задачи

$$C_{i\text{опт}} = \frac{\Lambda_i}{\mathcal{M}_i} + \sqrt{\frac{\Lambda_i}{\mathcal{M}_i s_i}} \sum_{j=1}^{m} \sqrt{\frac{\Lambda_i s_j}{\mathcal{M}_j}} \Bigg/ \gamma T^*, \tag{9.20}$$

$$\mathcal{G}_{\text{мин}} = \sum_{i=1}^{m} \frac{\Lambda_i s_i}{\mathcal{M}_i} + (\gamma T^*)^{-1} \left(\sum_{i=1}^{m} \sqrt{\frac{\Lambda_i s_i}{\mathcal{M}_i}}\right)^2, \tag{9.21}$$

где T^* — допустимое время задержки. Рассмотренные здесь оптимальные задачи поставил и решил Л. Клейнрок (Л. Клейнрок. Коммуникационные сети. — М.: Наука, 1970).

Подставляя значение $C_{i\,опт}$ из (9.18), (9.20) в (9.14) и округляя полученное значение до ближайшего целого числа, можно приближенно получить оптимальное число каналов в i-й МЛС.

Соотношения (9.18)—(9.21) позволяют для сети с коммутацией сообщений, заданной структурой и известными интенсивностями входящих потоков и характеристиками передачи сообщений в узлах приближенно выбирать оптимальное число каналов МЛС сети, чтобы при заданной арендной плате $\mathcal{G}^*$ средняя задержка сообщений в сети была минимальна. Если же задано допустимое среднее время задержки T^*, эти соотношения позволяют приближенно выбрать оптимальное число каналов МЛС, чтобы арендная плата за сеть была минимальна.

Для иллюстрации зависимостей $T(C_i)$, $\mathcal{G}(C_i)$ на рис. 9.1,*а*, *б* приведены графики, просчитанные на ЦВМ для одной из сетей. Штриховые линии указывают оптимальные значения $C_{i\,опт}$ и направления смещения экстремумов $T_{мин}$ и $\mathcal{G}_{мин}$. Анализ графиков показывает, что выбирая оптимальное значение пропускной способности МЛС, можно существенно повысить технико-экономическую эффективность сетей.

Контрольные вопросы

1. В чем сущность оптимизации пропускной способности уплотненных линий в сетях?
2. Сформулируйте и поясните прямую и обратную постановку задачи оптимизации пропускных способностей МЛС сети.
3. Поясните графики рис. 9.1.

9.5. УПЛОТНЕНИЕ ЛИНИЙ И ИНТЕГРАЛЬНЫЕ СЕТИ СВЯЗИ

На данном этапе развития техники связи невозможно рассматривать способы уплотнения линий без учета особенностей построения сетей связи и управления информационными потоками в них. Поэтому в теории сетей связи способы уплотнения и концентрации каналов рассматривают совместно. Особое внимание уделяется созданию интегральных сетей связи [4—6, 11], в которых используются общие принципы передачи дискретной (телеграфной и цифровой) и аналоговой (речевой) информации в цифровой форме и единые способы разделения каналов как при передаче сообщений по линиям, так и при их коммутации на автоматических телефонных станциях. Кратко рассмотрим общую характеристику способов уплотнения линий связи в интегральных сетях, покажем связь способов уплотнения с другими методами повышения эффективности сетей. Детально эти вопросы излагаются в курсе теории сетей связи.

Концентрация в теории сетей рассматривается как динамическая процедура распределения каналов, при которой n_1 входных каналов в соответствии с поступающими запросами на их использование динамически распределяется между меньшим количеством n_2 выходных каналов (обычно $n_1/n_2=(2—5)$). Концентрацию осуществляют с использованием коммутации каналов и комбинированных методов коммутации сообщений. В отличие от уплотнения концентрация предусматривает выравнивание нагрузки и повышение эффективности использования различных каналов линий связи. Скорости ввода и вывода информации в концентраторах не требуют согласования, но важную роль приобретают контроль и анализ статистической информации об изменении нагрузки и образовании очередей. Уплотнение и концентрация линий являются основными способами управления информационными потоками в сетях. Концентрацию с использованием коммутации каналов иногда называют пространственным уплотнением каналов.

Все большее распространение в сетях связи получают ИКМ и временное разделение каналов. Применение дискретных методов в сетях связи позволяет использовать все преимущества цифровой вычислительной техники для повышения эффективности связи: надежность, модульный принцип построения, использование интегральных схем, простота и др. А это в свою очередь способствует существенному снижению стоимости каналообразующей аппаратуры. Другим важным для построения сетей достоинством временного уплотнения каналов

является большое число каналов. Временное уплотнение хорошо совмещается с методами построения систем спутниковой связи, в которых предпочтение также отдается временному разделению. Например, магистральные линии сети передачи данных «*Datran*» имеют емкость 4000 каналов со скоростью передачи данных 4800 бит/с, модульная конструкция каналообразующего оборудования позволяет увеличить емкость до 8000 каналов. Применение ИКМ обеспечивает регенерацию сигналов в узлах связи, что исключает накопление шумов.

В сетях связи, где применяют ИКМ и временное разделение каналов, важную роль играет синхронизация. Аппаратуру временного уплотнения создают так, чтобы ошибки в каналах могли вызвать ошибки в передаваемых сообщениях, но не в сигналах синхронизации и управления, так как это ведет к значительным потерям информации. Это достигается без существенного снижения пропускной способности каналов связи, в результате применения корректирующего кодирования всех наиболее важных сигналов синхронизации и управления.

Следует отметить применение в сетях связи асинхронного временного уплотнения, которое является гибридной формой уплотнения и концентрации. Асинхронное временное уплотнение позволяет увеличить эффективность использования каналов связи благодаря тому, что каналы представляются пользователям только во время их активности. Для асинхронного временного уплотнения необходимо использовать и анализировать статистические данные по текущей активности абонентов, поэтому его часто называют статистическим уплотнением [8]. Недостатками этого вида уплотнения являются высокая стоимость значительно более сложных схем адресации, управления и накопления данных, возможность потерь информации при высоких нагрузках.

Таким образом, для построения интегральных сетей связи, предназначенных для передачи информации самого различного характера в цифровой форме, применяют методы синхронного и асинхронного временного уплотнения и методы концентрации, основанные на коммутации каналов, коммутации сообщений, а также на коммутации каналов и сообщений. Совместное рассмотрение этих методов с позиций анализа, синтеза и оптимизации сетей связи выполняется в теории сетей связи.

Контрольные вопросы

1. Какую роль играют способы уплотнения линий в интегральных сетях связи?
2. Что такое концентрация? Чем она отличается от уплотнения?
3. Какие способы уплотнения получили в сетях связи наибольшее распространение?
4. Какие меры принимают для повышения надежности синхронизации и управления в сетях связи?

9.6. ВЫВОДЫ

1. Основной проблемой уплотнения линий связи является разделение канальных сигналов — выделение канальных сигналов из искаженного принятого линейного сигнала. Необходимым условием разделимости является условие линейной независимости канальных сигналов (9.6), достаточным — условие их ортогональности (2.1).

2. Наиболее распространенными способами разделения канальных сигналов являются частотное, временное, частотно-временное разделение, разделение сигналов по форме и кодово-адресное разделение.

3. Пропускная способность уплотненной линии связи (9.10) определяется погрешностями формирования канальных сигналов, искажениями сигналов и помехами в линии связи, несовершенством разделяющих устройств приемников. Нельзя увеличить пропускную способность уплотненных линий, увеличивая мощность канальных сигналов, необходимо уменьшать и взаимное влияние каналов.

4. Рассмотрение уплотнения линий связи с более общих позиций повышения эффективности сетей связи, построенных на этих линиях, приводит к необходимости оптимального выбора пропускной способности (числа каналов) каждой

линии. Решают прямую и обратную задачу оптимизации пропускной способности уплотненных линий связи. В прямой задаче определяют оптимальную пропускную способность линии в сети связи, чтобы обеспечить среднее минимальное время задержки сообщений в сети при ограниченной арендной плате за пользование сетью. В обратной задаче минимизируют арендную плату за сеть при заданной допустимой средней задержке сообщений в сети. Оптимальная пропускная способность уплотненных линий связи сети растет с увеличением допустимой арендной платы за сеть, а минимальное среднее время задержки сообщений падает.

5. Уплотнение линий связи в сетях рассматривают совместно с другими способами управления информационными потоками и повышения эффективности передачи информации. Различные способы уплотнения играют важную роль в создании перспективных интегральных сетей связи, в которых все виды информации передаются в цифровой форме.

Наибольшее распространение в таких сетях получают импульсно-кодовая модуляция и временное уплотнение. В сочетании с методами концентрации (методами коммутации каналов и сообщений) методы цифровой передачи информации и различные способы временного уплотнения линий связи играют важную роль в повышении эффективности сетей. Рассмотрение методов уплотнения и концентрации с единых позиций производится в теории сетей связи.

Глава 10

ЭФФЕКТИВНОСТЬ ПЕРЕДАЧИ ИНФОРМАЦИИ

10.1. ОСОБЕННОСТИ ОЦЕНКИ ЭФФЕКТИВНОСТИ

Основной особенностью современного этапа развития теории информации и передачи сигналов является использование системного подхода к оценке эффективности передачи информации. При этом подходе учитываются все факторы, влияющие на достижение цели передачи информации и выполнение системой или сетью связи основных задач. Цель и задачи, решаемые системой, рассматривают с учетом ее влияния на другие системы, на окружающую среду и, особенно, на систему более высокого иерархического уровня. Можно определить одну из двух главных целей функционирования систем и сетей связи: обеспечение максимального количества передачи информации при фиксированных затратах и обеспечение требуемого качества передачи информации при минимальных затратах. И в том и в другом случае необходимо произвести оценку технико-экономической эффективности системы с учетом ее надежности и других факторов, влияющих на качество передачи информации [4, 8, 11, 16].

При выборе комплексного показателя технико-экономической эффективности системы исходят из того, что он должен иметь прямую связь с ее целевым назначением, объективно характеризовать все основные свойства, быть чувствительным к изменению определяющих параметров системы и наряду с этим должен быть достаточно простым, чтобы им можно было пользоваться на практике. Трудность заключается еще и в том, что не все цели системы можно адекватно отразить в количественной форме. Например, трудно измерить степень удовлетворения потребности людей в общении с помощью средств связи.

Разработка и исследование комплексных и интегральных показателей качества передачи информации — одна из важных и актуальных проблем. Ей уделяется большое внимание, и в этом направлении уже получены первые результаты. Частные показатели технической эффективности, которые нашли практическое применение в задачах оценки эффективности систем и сетей связи, рассматриваются в следующих параграфах.

Контрольные вопросы

1. Какой подход к оценке эффективности передачи информации используют на современном этапе развития теории информации и передачи сигналов?
2. Из каких соображений выбирают комплексные показатели эффективности систем и сетей связи?

10.2. ЭФФЕКТИВНОСТЬ ПЕРЕДАЧИ ДИСКРЕТНЫХ СООБЩЕНИЙ

10.2.1. Для оценки эффективности передачи дискретных сообщений применяют [2] *коэффициент использования мощности сигнала*

$$\beta_{\mathscr{P}} = \frac{R}{\mathscr{P}/\Omega_0} \text{ [бит]} \qquad (10.1)$$

и коэффициент использования полосы $\Delta F_к$ частот канала

$$\gamma_F = R/\Delta F_к \ [бит], \tag{10.2}$$

где R — скорость передачи информации; $\mathscr{P}$ — мощность сигнала; Ω_0 — спектральная плотность помехи. Учтем, что $\Omega_0 = \sigma^2/\Delta F_к$, $\mathscr{P}/\sigma^2 = h_1$, тогда

$$\beta_0 = \gamma_F / h_1. \tag{10.3}$$

Следовательно, показатели технической эффективности β и γ связаны простым соотношением (10.3), что позволяет пользоваться тем из них, который более подходит для оценки качества системы конкретного типа. Например, для оценки эффективности радиосистем чаще применяют показатель $_{\mathscr{P}}$, а для оценки эффективности проводных систем связи — γ_F.

По физическому смыслу $\beta_{\mathscr{P}0}$ и γ_F — это то удельное количество информации, которое можно передать по каналу при заданном отношении сигнал/шум или при фиксированной полосе канала.

Предельные оценки $\beta_{\mathscr{P}0}$ и γ_{F0} можно получить, используя в соотношениях (10.1)—(10.3) выражение для пропускной способности системы (5.38) вместо R; например, для двоичных систем, получим

$$\beta_{\mathscr{P}0} = [1 + p_0 \log p_0 + (1 - p_0) \log (1 - p_0)]/h_1, \tag{10.4}$$

$$\gamma_{F0} = [1 + p_0 \log p_0 + (1 - p_0) \log (1 - p_0)]/\Delta F_к T, \tag{10.5}$$

где $h_1 = PT/\Omega_0$; T — длительность сигнала.

Следовательно, предельное значение коэффициента использования мощности определяется вероятностью ошибки p_0 и отношением сигнал/шум. С ростом p_0 и h_1 величина β_0 уменьшается. Предельное значение коэффициента использования полосы канала падает с ростом p_0 и базы сигнала.

Для сравнения на рис. 10.1 [2] показаны графики зависимостей (10.5) (сплошные кривые) и (10.4) (штриховые кривые) от p_0 при фазовой телеграфии, частотной телеграфии и амплитудной манипуляции. Как и следовало ожидать, мощность наилучшим образом используется в системах ФТ, а полоса в системах ФТ и АМн. Минимальная ширина спектра высокочастотных сигналов при ФТ и АМн равна $2/T$, а ЧТ сигналов — $4/T$, поэтому кривые ФТ и АМн для γ_0 сливаются. Анализ графиков показывает, что повышение коэффициента использования мощности вызывает снижение коэффициента использования полосы частот. Рис. 10.1 наглядно иллюстрирует характер «обмена» мощности сигнала на полосу занимаемых частот.

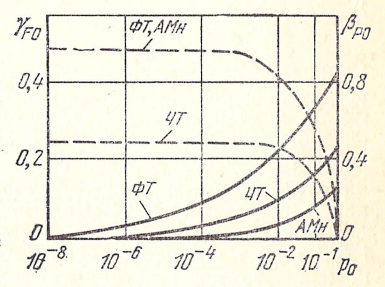


Рис. 10.1. Зависимость $\beta_{\mathscr{P}0}$ и γ_{F0} от p_0

10.2.2. Для повышения эффективности передачи дискретных сообщений применяют разнесенный прием сигналов, прием в целом, каналы обратной связи, широкополосные (шумоподобные) сигналы, адаптивную коррекцию характеристик каналов и другие способы.

Сущность *разнесенного приема* рассмотрена в § 6.6. Разнесенный прием существенно повышает эффективность передачи дискретной информации, так как снижает вероятность ошибки. Вероятность ошибки разнесенного приема по ме-

тоду голосования, когда решение принимается по большинству принятых нулей или единиц,

$$p_n = \sum_{i=k}^{n} C^i{}_n p^i{}_0 (1 - p_0)^{n-i}, \tag{10.6}$$

где k — число каналов, в которых произошла ошибка; n — число независимых каналов. Обычно n выбирают нечетным, чтобы упростить схему голосования.

Предположим, что $p_0 = 10^{-3}$, $n = 3$, тогда $p_3 \simeq 3p^2{}_0 \simeq 3 \cdot 10^{-6}$. Следовательно, даже в постейшем случае разнесенный прием приводит к уменьшению вероятности ошибки примерно на три порядка.

Сущность *приема в целом* заключается в том, что вместо двух решающих схем в приемнике применяют одну и решение принимают по сигналу, соответствующему всей кодовой комбинации. Эффективность приема в целом выше благодаря тому, что в работе приемника используется та часть полезной информации с выхода первой решающей схемы, которая при поэлементном приеме теряется после того, как первая решающая схема приняла решение. Поэлементный прием и прием в целом равноценны только в случае кодирования без избыточности.

Аппаратурная реализация приема в целом достаточно сложна и требует применения набора фильтров, согласованных с сигналами, которые соответствуют всем разрешенным кодовым комбинациям. Поэтому разрабатывают методы приема, промежуточные между поэлементным приемом и приемом в целом, в которых во второй решающей схеме используется часть информации о непрерывном сигнале на входе первой решающей схемы. Эффективность передачи сообщений такими методами возрастает, а сложность аппаратурной реализации остается приемлемой.

Системы с каналами обратной связи являются разновидностью *дуплексных систем* (системы, в которых может производиться одновременно независимый двусторонний обмен информацией между пользователями), в которых пропускную способность каналов обоих направлений частично используют для передачи дополнительной информации, способствующей повышению эффективности передачи. Обычно по каналу обратной связи передают лишь сигналы, полезные для коррекции ошибок в прямом канале. Различают системы с информационной обратной связью (ИОС) и системы с решающей (управляющей) обратной связью (РОС).

В системах с ИОС по каналу обратной связи непрерывно поступает информация о том, в каком виде по прямому каналу принимаются сообщения. По результатам анализа этой информации передающее устройство принимает решение о степени соответствия принятого сообщения переданному и изменяет способ передачи, если имеются ошибки. Например, может быть увеличена мощность передатчика, выполнено полное повторение ошибочно принятого сообщения, изменена избыточность кода, полностью прекращена передача, если прямой канал находится в плохом состоянии.

Если вероятность ошибки в обратном канале близка к нулю, применение ИОС может обеспечить существенное повышение эффективности передачи сообщений. Это достигается ценой усложнения аппаратуры и снижения скорости передачи информации, так как по существу используется временная избыточность для повышения верности.

В системах с РОС в приемнике непрерывно осуществляется оценка верности передачи сообщений и лишь при появлении ошибки по каналу обратной связи посылается сигнал запроса повторения. Обратный канал используется значительно реже, чем в системах с ИОС. В остальное время он может применяться как прямой канал обратного направления. Это существенное достоинство систем с РОС. Использование корректирующих кодов и детекторов качества сигналов позволяет существенно повысить верность и скорость передачи. Применение обратной связи упрощает коды, так как они работают в режиме обнаружения ошибок. Соответственно упрощается и аппаратурная реализация систем.

Применение шумоподобных переносчиков позволяет не только приблизить скорость передачи информации к пропускной способности каналов, но и сущест-

венно повысить верность передачи сообщений, особенно по многолучевым радиоканалам с сосредоточенными помехами. База шумоподобных переносчиков $\nu = 10^2$—10^3. Используя многолучевое распространение как основу разнесенного приема, удается существенно повысить верность и в радиоканалах с селективными замираниями.

В широкополосных системах интервал корреляции коэффициента передачи канала гораздо меньше длительности сигналов. Если время взаимного запаздывания сигналов будет больше времени их корреляции, то сигналы отдельных лучей будут полностью разделены коррелятором. Тогда для использования разнесенного приема остается выполнить коррекцию их фаз и амплитуд, а затем сформировать результирующий сигнал. Повышение помехоустойчивости достигается введением временной и частотной избыточности сигналов.

В одной из первых широкополосных КВ радиостанций «Rake» (1956 г.) широкополосные сигналы создавались на основе линейных рекуррентных последовательностей биполярных коротких импульсов длительностью $\tau = 1/F$, расположенных псевдослучайно внутри сигнала общей длительностью $T \gg \tau$ (см. § 3.3). В системе «Rake» $F = 10$ кГц, $T = 22 \cdot 10^{-3}$ с, поэтому $\nu/2 = 220$. В настоящее время сигналы создаются с помощью регистров сдвига, охваченных цепями обратной связи (см. § 8.2). Обеспечение верности передачи достигается за счет ухудшения эффективности использования полосы частот, так как $\gamma_0 \simeq 2/\nu \ll 1$ (см. (10.5)).

Адаптивная коррекция характеристик каналов заключается в периодическом зондировании канала испытательным импульсом, который отделен от информационного пакета сигналов защитным интервалом, в текущем измерении и, если требуется, в коррекции коэффициента передачи канала. Синхронные системы адаптивной коррекции каналов сочетают возможности обеспечения высокой скорости передачи дискретных сообщений и временного уплотнения линий связи.

Контрольные вопросы

1. Какие применяют показатели технической эффективности передачи дискретных сообщений?
2. Сравните эффективность систем с ФТ, ЧТ и АМн двоичными сигналами.
3. Поясните сущность способов повышения эффективности передачи дискретных сообщений.

10.3. ЭФФЕКТИВНОСТЬ ПЕРЕДАЧИ НЕПРЕРЫВНЫХ СООБЩЕНИЙ

10.3.1. Оценка эффективности. Уже отмечалось, что оценка эффективности передачи непрерывных сообщений — это оценка эффективности способа модуляции. Для оценки эффективности обычно используют выигрыш по отношению сигнал/шум и коэффициент использования пропускной способности канала (5.87)

$$\eta_1 = [\log(1 + \mathfrak{y} h_1)]/\nu_1 \log(1 + h_1).$$

Для оценки эффективности $\mathfrak{y}$ многоступенчатых и цифровых видов модуляции применяют методы, изложенные в § 7.5, 7.6.

В табл. 10 [2, 3] приведены данные сравнительного анализа эффективности различных видов модуляции, полученные при $\mathfrak{y} h_1 \simeq 40$ дБ и пик-факторе $\Pi = \sqrt{2}$ для гауссова канала при оптимальной обработке сигналов. Анализ показывает, что лучше всего используется пропускная способность канала при однополосной модуляции, однако потенциальная помехоустойчивость этого вида модуляции низка ($\eta_1 = 1$). Фазовая и частотная модуляция при больших индексах приближаются по потенциальной помехоустойчивости к идеальной модуляции (ИС) (выигрыш составляет десятки и сотни раз), но коэффициент использования канала мал (0,19—0,32) из-за большой частотной избыточности модулированных сигналов.

Таблица 10

Метод модуляции	ν_1	$\mathfrak{y}$	$\mathfrak{y}_1$	η_1
АМ	2	0,67	0,36	0,48
БМ	2	1	0,5	0,5
ОМ	1	1	1	1
ФМ	2	1	0,5	0,5
ФМ	20	1000	50	0,19
ЧМ	2	3	1,5	0,5
ЧМ	20	3000	150	0,32
ИС	20	6310	315	1

10.3.2. Повышение эффективности передачи непрерывных сообщений. Основными способами повышения эффективности передачи непрерывных сообщений являются устранение избыточности, статистическое уплотнение, применение цифровых видов модуляции.

Большая *избыточность непрерывных сообщений* — одна из основных причин снижения эффективности систем. Поэтому развиваются способы *устранения избыточности.* Все они основаны на дискретной передаче не самого сигнала, а его наиболее характерных параметров, изменение которых во времени протекает гораздо медленнее, чем изменение самого сигнала. Этот принцип используют в вокодерах, видеокодерах и телекодерах — устройствах для компрессии спектров телефонных, видеотелефонных и телевизионных сигналов.

Вокодеры (сокращение английского названия «voice coder» — кодировщик голоса) — это телефонные системы с параметрической компрессией спектров речевых сигналов. Вместо текущего спектра передается информация о параметрах сигналов: спектральных уровнях в узких диапазонах частот сигнала, частотах и уровнях формант (областей спектра сигнала с увеличенным значением амплитуд), основном тоне, фонемах (элементах речи, воспринимаемых как одно целое и однородное; для речи фонемы то же, что буквы для письма) и др. Избыточными элементами речевых сигналов считают высоту тона, тембр, акцент, интонацию и т. п. На рис. 10.2 показана структурная схема вокодера. Анализатор сообщений выделяет наиболее характерные параметры речевого сигнала. Формирователь тем или иным способом кодирует процесс изменения этих параметров для передачи по линии связи. В приемнике при декодировании выполняется оценка принятых параметров сигнала и полученные оценки используются в синтезаторе сообщений для воспроизведения копии речевого сигнала.

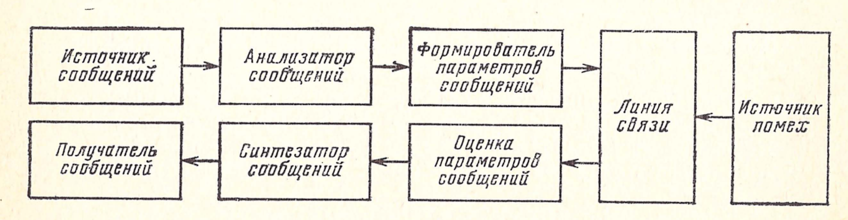


Рис. 10.2. Обобщенная структурная схема вокодера

Теоретически спектр речевого сигнала можно сжать примерно в 200 раз, что позволяет использовать полосу 15—20 Гц. Однако предельная компрессия аппаратурно трудно реализуема. Кроме того, теряется узнаваемость речи говорящего. Разработаны вокодеры с сохранением узнаваемости, которые используют полосу 100—200 Гц, что соответствует 15—30-кратной компрессии спектра и такому же увеличению дополнительного числа каналов в уплотненных линиях. Например, один из действующих малогабаритных цифровых вокодеров занимает объем 0,12 м^3, включая и источники питания, и обеспечивает высокую натуральность речи при скорости передачи информации 2400 бит/с [8].

Задачи устранения избыточности видеотелефонных и телевизионных сообщений еще более актуальны. Достаточно указать, что только двукратная компрес-

сия спектра телевизионного сигнала позволяет организовать в освободившейся полосе частот в 2 МГц дополнительно около 700 телефонных каналов. Передача видеотелефонного сигнала в аналоговой форме требует полосы около 1 МГц. Если применить цифровую передачу методом дифференциальной ИКМ, использовать статистическое кодирование, линейное и нелинейное предсказание сигналов, можно обеспечить более чем 5—10-кратную компрессию спектра видеотелефонного сигнала. Например, цифровой видеокодер фирмы «Филко — Форд» для передачи черно-белых малоподвижных изображений (говорящих по телефону людей) использует стандартный телефонный канал [8]. Успехи в создании телекодеров пока еще не столь значительны.

В обычных системах статистические особенности использования каналов абонентами не учитывают. Если эти особенности учесть, можно существенно повысить эффективность передачи сообщений. Сущность *статистического уплотнения* заключается в том, что паузы в передаче и свободные полосы частот, обусловленные статистическими особенностями источников информации, используют для организации дополнительных каналов и передачи дополнительной информации [8].

Наибольшее распространение получили телефонные системы статистического уплотнения, в которых дополнительная аналоговая или дискретная информация передается в паузах между речевыми сигналами. Передача дополнительной информации повышает коэффициент использования уплотненных линий с 37 до 90% [8]. Эффективность статистического уплотнения тем выше, чем больше каналов. Например, в 12-канальной системе эффективность статистического уплотнения равна 1,5 (в среднем образуется дополнительных 6 каналов), в 96-канальной системе эффективность статистического уплотнения 2,45, следовательно, дополнительно организуется около 140 каналов — больше, чем в самой системе.

Управление динамической нагрузкой каналов позволяет повысить эффективность статистического уплотнения до 2,74 раза. Сущность этого способа заключается в контроле и выравнивании реальной нагрузки различных каналов, в результате коэффициенты использования всех каналов примерно одинаковы.

Сейчас уже очевидно, что *применение цифровых способов* для повышения эффективности передачи непрерывных сообщений является магистральным направлением. Развитие интегральной микроэлектроники и цифровой вычислительной техники, проникновение в технику связи программных способов управления процессами передачи сообщений, преимущества унификации и стандартизации цифровых элементов и модулей аппаратуры, приближение характеристик цифровых видов модуляции к характеристикам идеальной модуляции — все это создает объективные предпосылки для дальнейшего развития высокоэффективных цифровых способов передачи непрерывных сигналов.

Контрольные вопросы

1. Какие применяют показатели технической эффективности передачи непрерывных сообщений?
2. Сравните эффективность различных видов модуляции.
3. Поясните сущность и сравните известные способы повышения эффективности передачи непрерывных сообщений.

10.4. ЭФФЕКТИВНОСТЬ ПЕРЕДАЧИ ИНФОРМАЦИИ В СЕТЯХ

Современные сети связи построены с использованием большого числа разнородных систем, узлов и линий связи. Поэтому для оценки и повышения эффективности передачи информации в сетях находят применение все рассмотренные ранее методы. Однако этого недостаточно, так как сети связи являются качественно новым видом объектов информационной техники, что ведет к необходимости разработки принципиально новых методов оценки и повышения эффективности [11]. Детально все эти методы изучают в курсе «Теория сетей связи», мы рассмотрим лишь характерные особенности оценки эффективности сетей и способов ее повышения.

В отличие от систем связи конкретного назначения, для которых скорость передачи информации и достоверность являются основными техническими характеристиками качества, в сетях связи на первый план выступают такие комплексные показатели эффективности передачи информации в сетях, как общее число узлов или конечных пунктов, общая длина линий (каналов), годовой объем переданной информации, стоимость сети или приведенные затраты, номинальная мощность сети по пропускной способности, фактическая загрузка сети, время доставки информации, относительное число потерянных вызывов и задержек, показатели полезного использования времени абонентов, коэффициенты готовности основных элементов сети [16], показатели связности и живучести, показатели использования узлов и каналов, и многие др. [11].

Особенности выбора и использования для оптимизации сети таких критериев эффективности, как среднее время задержки сообщений и арендная плата за сеть, рассмотрены в § 9.4. В любой сети изменение скорости передачи информации, увеличение времени установления соединения, увеличение уровня помех в каналах и искажение сигналов, отказы аппаратуры и другие процессы неизбежно влияют на задержку сообщений в сети. Поэтому среднее время задержки сообщений в сети можно рассматривать как комплексный критерий.

Для повышения эффективности передачи информации в сетях могут быть использованы любые технически и экономически целесообразные способы. В числе основных способов следует отметить переход на автоматическую коммутацию каналов и сообщений, оптимальное сочетание программных и аппаратурных методов коммутации и управления, применение цифровых способов передачи информации, коммутации и управления (интеграция принципов работы систем сети), повышение качества синхронизации и применение асинхронного принципа, использование все более высоких частот радиоканалов (развитие радиорелейных линий), применение спутниковых линий связи для резервирования наземных линий, оптимальное сочетание методов централизованного и децентрализованного управления сетями, применение динамического согласования нагрузки различных линий и каналов сетей, разработку теоретических методов структурной и параметрической оптимизации сетей с многополюсными потоками, многоадресными сообщениями и с приоритетами сообщений, все более широкое внедрение элементов и устройств цифровой вычислительной техники.

Контрольные вопросы

1. Назовите комплексные показатели эффективности передачи информации в сетях.
2. Какие используют способы повышения эффективности передачи информации в сетях?

10.5. ВЫВОДЫ

1. Оценка эффективности передачи информации была и остается актуальной задачей теории информации и передачи сигналов. По мере развития теории и техники связи все более очевидной становится необходимость системного подхода к оценке эффективности. Системный подход позволяет произвести технико-экономическую оценку эффективности с учетом взаимосвязи основных технических и экономических показателей качества передачи информации. Это направление интенсивно развивается.

2. Для оценки эффективности передачи дискретных сообщений практическое распространение получили коэффициент использования мощности сигнала и коэффициент использования полосы частот канала. По физическому смыслу оба коэффициента показывают то максимальное удельное количество дискретной информации, которое можно передать по каналу при фиксированном отношении сигнал/шум или при фиксированной полосе частот канала. Увеличение эффективности использования мощности всегда связано с уменьшением коэффициента использования полосы.

Эффективность передачи дискретных сообщений можно повысить, применяя разнесенный прием сигналов, прием сигналов в целом, широкополосные сигналы,

адаптивную коррекцию характеристик каналов связи, вводя каналы обратной связи, и другими методами.

3. Оценка эффективности передачи непрерывных сообщений сводится к оценке эффективности метода модуляции. Для этого обычно используют обобщенный выигрыш по отношению сигнал/шум и коэффициент использования пропускной способности канала. При цифровой передаче непрерывных сообщений используют методы оценки эффективности многоступенчатых видов модуляции. Эффективность передачи непрерывных сообщений повышают, устраняя избыточность сообщений, применяя статистическое уплотнение, цифровые способы передачи и др.

4. Оценка эффективности передачи информации в сетях является сложной и малоизученной проблемой. Особенности выбора и использования таких критериев эффективности, как среднее время задержки сообщений и арендная плата за сеть рассмотрены в § 9.4. Эти комплексные критерии связаны со всеми основными техническими и экономическими показателями качества передачи информации и могут быть полезны при решении задач анализа эффективности и оптимального синтеза сетей.

Для повышения эффективности передачи информации в сетях применяют большинство рассмотренных методов теории информации и передачи сигналов, а также разрабатывают принципиально новые методы, которые в совокупности образуют теорию сетей связи.

ЗАКЛЮЧЕНИЕ

Рождение теории информации и передачи сигналов как самостоятельной науки относят к середине 50-х годов и связывают с фундаментальными работами В. А. Котельникова и К. Шеннона. За тридцать лет эта наука развилась в такой степени, что стала учебной дисциплиной во многих вузах нашей страны.

Трудами советских и зарубежных ученых и инженеров созданы и продолжают создаваться фундаментальные методы анализа, синтеза и оптимизации информационных систем. Развивается математическое моделирование процессов передачи сообщений, разрабатываются и внедряются новые методы модуляции, методы анализа нестационарных непрерывных каналов, неоднородных асимметричных дискретных каналов, неоднородных асимметричных дискретных каналов с памятью, развиваются теория многолучевых радиоканалов, помехоустойчивые методы и алгоритмы передачи информации, в которых учитываются ненадежность аппаратуры и отклонение характеристик устройств от идеальных, строятся математические модели сложных помеховых ситуаций, интенсивно развивается корректирующее кодирование, дальнейшее развитие получают теория систем с обратной связью, теория адаптивных систем [12] и т. д. и т. п.

Методы теории информации и передачи сигналов все шире применяют для решения практических задач повышения качества передачи информации и эффективности систем и сетей связи, радионавигационных и радиолокационных систем, автоматизированных систем управления воздушным движением и посадки самолетов, вычислительных систем, информационно-логических измерительных комплексов и многих других.

Основные тенденции и перспективы теории информации и передачи сигналов следующие:

— широкое внедрение цифровых методов передачи сообщений и временного уплотнения каналов и использование в теории и технике связи методов и средств цифровой вычислительной техники;

— рост удельного веса алгоритмических и программных методов управления процессами передачи информации;

— применение широкополосных методов и систем передачи информации;

— широкое использование детекторов качества сигналов и корректирующего кодирования;

— применение адаптации для оперативной коррекции характеристик каналов, устранения избыточности сообщений, статистического уплотнения и т. п.;

— разработка методов оценки эффективности передачи информации с позиций системного подхода;

— применение цифрового и статистического моделирования процессов передачи сообщений для анализа и синтеза информационных систем;

— широкое внедрение функционально-модульного принципа построения информационных систем.

Глубокое и всестороннее развитие теории информации и передачи сигналов, тесно связанное с практическими потребностями информационной техники, творческое освоение ее методов учеными и инженерами позволяет успешно повышать качество передачи информации и эффективность эксплуатируемых и проектируемых систем.

ПРИЛОЖЕНИЕ

ОСНОВНЫЕ ПОНЯТИЯ ФУНКЦИОНАЛЬНОГО АНАЛИЗА

В теории информации и передачи сигналов широко используют такие фундаментальные понятия функционального анализа, как преобразование пространств, функция, оператор, функционал, функциональный оператор, линейность и нелинейность преобразования и др. Поэтому необходимо в систематизированном и доступном виде кратко изложить основные сведения об этих понятиях.

Рассмотрим два множества математических объектов **A** и **B** (чисел, функций и т. п.) и предположим, что нам известен закон Q, согласно которому любому объекту A_i, принадлежащему множеству **A** ($A_i \in \mathbf{A}$), ставится в соответствие определенный объект $B_j \in \mathbf{B}$.

Преобразование **A** в **B** можно отразить формулой

$$B_j = Q(A_i). \qquad (\text{П.1})$$

Это преобразование будет обладать свойством линейности, если выполняются условия аддитивности

$$Q(A_1 + A_2) = Q(A_1) + Q(A_2) \qquad (\text{П.2})$$

и однородности

$$Q(\lambda A_i) = \lambda Q(A_i), \qquad (\text{П.3})$$

где λ — любое произвольное число.

Если объекты **A** и **B** — это числа x и y, то закон Q — это функция

$$y = Q(x), \qquad (\text{П.4})$$

которая устанавливает соответствие между числами x и y.

Свойством линейности обладает линейная функция $y = kx$, где k — известное число.

Если объекты **A** и **B** — это функции одной и той же переменной, например сигналы $s_1(t)$ и $s_2(t)$, то закон Q — это оператор, который указывает способ преобразования $s_1(t)$ в $s_2(t)$:

$$s_2(t) = Q[s_1(t)]. \qquad (\text{П.5})$$

Свойством линейности обладает линейный оператор. Примерами линейных операторов служат операторы интегрирования и дифференцирования:

$$\frac{d}{dt}[\lambda_1 s_1(t) + \lambda_2 s_2(t)] = \lambda_1 \frac{ds_1(t)}{dt} + \lambda_2 \frac{ds_2(t)}{dt}, \qquad (\text{П.6})$$

$$\int [\lambda_1 s_1(t) + \lambda_2 s_2(t)]dt = \lambda_1 \int s_1(t)\, dt + \lambda_2 \int s_2(t)\, dt. \qquad (\text{П.7})$$

Примерами нелинейных операторов служат операторы логарифмирования и возведения в степень:

$$\ln[\lambda_1 s_1(t) + \lambda_2 s_2(t)] \neq \ln[\lambda_1 s_1(t)] + \ln[\lambda_2 s_2(t)], \qquad (\text{П.8})$$

$$[\lambda_1 s_1(t) + \lambda_2 s_2(t)]^2 \neq \lambda^2{}_1 s^2{}_1(t) + \lambda^2{}_2 s^2{}_2(t). \qquad (\text{П.9})$$

Если элемент A_i множества **A** — это функция $f(x)$, элемент B_j множества **B** — это число z, то преобразование Q — это функционал

$$z = Q[f(x)]. \qquad (\text{П.10})$$

Функциональный оператор — это распространение понятия функционала на более общий случай, когда элементом множества **A** является функция одной пе-

ременной $f(x)$, а элементом B_j множества **B** — функция другой переменной $\varphi(y)$. Тогда

$$\varphi(y) = Q[f(x)], \tag{П.11}$$

т. е. функциональный оператор устанавливает взаимное и однозначное соответствие функции одной переменной и функции другой переменной. Примерами функциональных операторов являются преобразования Фурье, Хинчина — Винера (2.22). Функциональный оператор можно рассматривать как одну из форм функционала потому, что при каждом фиксированном y число $\varphi(y)$ является функционалом $f(x)$.

Укажем характерные примеры аппаратурной реализации основных преобразований. Генератор развертки в осциллографе реализует операцию (П.4), он создает напряжение, которое является линейной функцией времени. Усилитель реализует (П.5), выходной сигнал усилителя пропорционален входному сигналу. Интегратор реализует (П.10), выходной сигнал интегратора пропорционален определенному интегралу от входного сигнала. Анализатор спектра реализует (П.11), входным сигналом анализатора является напряжение как функция времени, а выходным — спектральная плотность как функция частоты.

СПИСОК ЛИТЕРАТУРЫ

1. **Кловский Д. Д.** Теория передачи сигналов: Учебник для вузов. — М.: Связь, 1973. — 376 с.
2. **Зюко А. Г., Коробов Ю. Ф.** Теория передачи сигналов: Учебник для вузов.— М.: Связь, 1972. — 282 с.
3. **Назаров М. В., Кувшинов Б. И., Попов О. В.** Теория передачи сигналов: Учебник для вузов. — М.: Связь, 1970. — 386 с.
4. **Передача** дискретной информации и телеграфия: Учебник для вузов/ В. С. Гуров, Г. А. Емельянов, Н. П. Етрухин, В. Г. Осипов. — М.: Связь, 1974. — 526 с.
5. **Пенин П. И.** Системы передачи цифровой информации: Учебное пособие для вузов. — М.: Сов. радио, 1976. — 386 с.
6. **Цапенко М. П.** Измерительные информационные системы: Учебное пособие для вузов. — М.: Энергия, 1974. — 320 с.
7. **Зиновьев А. Л., Филиппов Л. И.** Введение в теорию сигналов и цепей: Учебное пособие для вузов. — М.: Высшая школа, 1975.—264 с.
8. **Босый Н. Д., Игнатов В. А.** Многоканальные системы передачи информации. — М.: Знание, 1974. — 64 с. — (Радиоэлектроника).
9. **Финк Л. М.** Теория передачи дискретных сообщений. — М.: Сов. радио, 1970. — 728 с.
10. **Элементы** теории передачи дискретной информации/ Л. П. Пуртов, А. С. Замрий, А. И. Захаров, В. М. Охорзин; Под ред. Л. П. Пуртова. — М.: Связь, 1972. — 232 с.
11. **Давыдов Г. Б., Рогинский В. Н., Толчан А. Я.** Сети электросвязи. — М.: Связь, 1977. — 360 с.
12. **Репин В. Г., Тартаковский Г. П.** Статистический синтез при априорной неопределенности и адаптации информационных систем. — М.: Сов. радио, 1977. — 432 с.
13. **Ван Трис Г.** Теория обнаружения, оценок и модуляции. В 3-х т.: Пер. с англ./ Под ред. В. И. Тихонова. — М.: Сов. радио, 1972. — Т 1. — 744 с.
14. **Свириденко С. С.** Основы синхронизации при приеме дискретных сигналов. — М.: Связь, 1974.—144 с.
15. **Левин Б. Р.** Теоретические основы статистической радиотехники. В 3-х кн. — М.: Сов. радио, 1974. — Кн. 1. — 2-е изд. — 752 с. — 1975. — Кн. 2. — 2 изд. — 504 с. — 1976. — Кн. 3. — 288 с.
16. **Игнатов В. А., Маньшин Г. Г., Трайнев В. А.** Статистическая оптимизация качества функционирования электронных систем. — М.: Энергия, 1974. — 264 с.
17. **Словарь** терминов по информатике/ Г. С. Жданова, Е. С. Колобродова, В. А. Полушкин и др.; Под ред. А. И. Михайлова. — М.: Наука, 1971. — 360 с.
18. **Энциклопедия кибернетики.** В 2-х т. Под ред. В. М. Глушкова. — Киев.: Гл. ред. Украинской Советской энциклопедии, 1974. — Т. 1. — 608 с. — Т. 2 — 620 с.

УСЛОВНЫЕ ОБОЗНАЧЕНИЯ

$A(t)$ — оригинал сообщения
$B(t)$ — копия сообщения
$s(t)$ — передаваемые модулированные сигналы
$z(t)$ — принятый сигнал
$\xi(t)$ — помеха (шум)
K — оператор кодирования
L — оператор преобразования сигнала в линии связи
$D_к$ — оператор декодирования
M_1 — оператор модуляции
$D_м$ — оператор демодуляции
a_i — i-й символ дискретного сообщения
K_i — i-я кодовая комбинация
T — длительность сигнала
ΔF_c — ширина спектра сигнала
Д — динамический диапазон сигнала (канала)
V — объем сигнала (канала)
$\nu = 2T\Delta F$ — база сигнала, (канала)
$x(t)$ — сигнал-переносчик
$u(t)$ — полезный сигнал
m — объем алфавита (основание кода)
n — длина сообщения (значность кода)
N — общее число сообщений (кодовых комбинаций)
I — количество информации
H — энтропия источника
r — коэффициент избыточности сообщений (кода)
R — скорость передачи информации
W — скорость передачи сигналов
C — пропускная способность
p_0 — вероятность ошибки при передаче элемента сигнала
$P(a_{i1}|a_{i2})$ — условная апостериорная вероятность правильного приема
Ω_0 — спектральная плотность белого шума
d — расстояние
$\oplus$ — символ сложения по модулю
$<\cdot>$ — усреднение по времени
$M[\cdot]$ — усреднение по ансамблю
$s_1 s_2$ — скалярное произведение двух векторов
$f_1(\cdot)$ — одномерная плотность распределения
$f_n(\cdot)$ — п-мерная плотность распределения
$f(z|s_i)$ — функция правдоподобия
E — энергия сигнала (помехи)
$\mathcal{P}$ — мощность сигнала (помехи)

ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

ОГЛАВЛЕНИЕ

Глава 4.

Глава 5.

Глава 6.

Глава 7.

Глава 8.

Глава 9.

Глава 10.

Игнатов В. А.

И 26 Теория информации и передачи сигналов: Учебник для вузов. — М.: Сов. радио, 1979. — 280 с., ил.

В пер.: 85 к.

Систематизированно излагаются основные положения теории информации и передачи сигналов. Излагаются методы математического описания сообщений, сигналов, помех и каналов связи, методы управления информационными параметрами сигналов, информационные характеристики источников сообщений, сигналов, помех и каналов, помехоустойчивость передачи дискретных и непрерывных сообщений, корректирующее кодирование. Рассмотрены принципы уплотнения линий связи, оценки и повышения эффективности передачи информационных потоков в сетях.

Книга является учебником для студентов радиотехнических факультетов вузов гражданской авиации.

И $\frac{30401\text{-}047}{046(01)\text{-}79}$ 47-79 1502000000

ББК 32.81

6Ф1

ИБ № 503

ВЛАДИМИР АЛЕКСЕЕВИЧ ИГНАТОВ

Теория информации и передачи сигналов

Редактор *Е. В. Вязова*
Художественный редактор *А. Н. Алтунин*
Художник *В. В. Волков*
Технические редакторы *В. А. Позднякова, И. В. Орлова*
Корректор *Л. А. Максимова*

Сдано в набор 16.10.78 Подписано в печать 16.04.79 Т-07057
Формат 60×90/16 Бумага типографская № 2 Гарнитура литерат.
Печать высокая. Объем 17,5 усл. п. л., 19,6 уч.-изд. л.
Тираж 15 000 экз. Заказ 886 Цена 85 к.

Издательство «Советское радио», Москва, Главпочтамт, а/я 693

Московская типография № 10 «Союзполиграфпрома»
Государственного Комитета СССР
по делам издательств, полиграфии и книжной торговли.
Москва, М-114, Шлюзовая наб., 10.

85 к.



В. А. ИГНАТОВ • ТЕОРИЯ ИНФОРМАЦИИ И ПЕРЕДАЧИ СИГНАЛОВ